10/048154
JC13 Rec'd PCT/PTO 29 JAN 2002

DESCRIPTION

STOMACH CANCER-ASSOCIATED GENE

Technical Field

The present invention relates to stomach cancer-associated genes.

Background Art

Stomach cancer is a cancer especially observed among Japanese in the world, and is a considerable disease that ranks high as the cause of death due to cancer in Japan. Excellent results, a 5-year survival rate exceeding 90%, have been obtained in the cases where the stomach cancer were detected early and could be surgically treated at an early stage. On the other hand, in cases of inoperable progressive cancers and cancers with metastasis, prognosis is unsatisfactory due to the fact that effective anti-cancer drugs have not been developed.

The lack of development of tumor markers specific for stomach cancer, which will be clinically useful, makes the detection of stomach cancer at early stages difficult. There are only few reports on genes which expression increases in relation with carcinogenesis and malignant alteration of stomach cancer, and thus, no indicator for stomach cancer leading to the detection at an early stage has been known. Therefore, indirect radiography has been widely used as a screening method aiming to detect stomach cancer at an early stage. However, increase in the opportunity of exposure to X ray, the fact that the examination results are dramatically influenced by roentgenography reading techniques, and such have been pointed out as problems. Later on, it was reported that the value of the serum pepsinogen reflects the atrophic gastritis, a leading lesion of stomach cancer, which has been applied to methods for screening stomach cancer. However, pepsinogen is a precursor of digestive enzyme secreted in stomach, and thus, cannot be used as a target molecule for the treatment of stomach cancer. Moreover, the pepsinogen method cannot be used as an indicator for the malignancy level of stomach cancer.

Identification of a causative gene for stomach cancer may enable the detection of stomach cancer at an early stage using the expression

level and activation of the gene as an indicator. Alternatively, it is also expected that detection of stomach cancer at an early stage and prognostic estimation may be facilitated by discovering a gene which expression level changes in relation with carcinogenesis and malignant alteration of stomach cancer.

There are also cases often observed where the cancer did not heal in spite of the surgical excision of the primary focus among patients with stomach cancer. The main cause of this fact is peritoneal metastasis (Gekachiryo 75: 96-102, 1996; Jpn. Surgery 19: 153, 1989). Peritoneal metastasis is most frequently observed among postoperative recurrence types of stomach cancer. Various methods for treating peritoneal metastasis have been attempted, however the results are far from sufficient. Peritoneal metastasis is a progressive form characteristic of scirrhous stomach cancer (Jpn J. Pathology 81: 21-49, 1992).

Peritoneal metastasis of stomach cancer is assumed to be formed through a simple process comprising implantation and proliferation of cancer cells released from the serosa on the peritoneum. However, not all the cancer cells released into the peritoneal cavity results in the formation of metastasis. This can be also predicted from the low frequency of metastasis formation after transplantation of cells derived from scirrhous stomach cancer into the peritoneal cavity of a nude mouse. Therefore, it is inferred that only cells having particular characters may result in metastasis formation, but the detailed mechanism of metastasis formation have not been elucidated yet.

According to previous reports to date, peritoneal metastasis has been detected comparatively more in scirrhous stomach cancer with the following characteristics (Jpn J. Gastroenterological Surgery 23: 1813-1820, 1990; Jpn J. Gastroenterological Surgery 25: 763-774, 1992):

macroscopic type: invasive type 3 or type 4,
histological type: low differentiation type, and
high lymph node metastasis positive case.

However, it is actually difficult to describe the peritoneal metastatic property based only on these clinical pathological

characteristics. Therefore, to elucidate the mechanisms of peritoneal metastasis, a high peritoneal matastatic cell strain, OCUM-2MD3, has been established, which is a substrain derived from parent strain OCUM-2M that hardly causes peritoneal metastasis. The parent strain OCUM-2M is a stomach cancer cell strain established from a scirrhous stomach cancer focus, which scarcely causes peritoneal metastasis even if it is inoculated into the peritoneal cavity of a nude mouse. The substrain OCUM-2MD3, on the other hand, causes formation of peritoneal metastasis by 100% by inoculating 5×10^6 cells or more (Br. J. Cancer 72: 1200-1210, 1995; Clin & Exp. Metastasis 14: 43-54, 1996). The OCUM-2MD3 is a cell strain established by inoculating the parent strain OCUM-2M into the peritoneal cavity of mice, collecting cells which caused peritoneal metastasis, reproliferating the cells in a culture system, inoculating these cells into the peritoneal cavity of nude mice, and isolating cells from the detected metastatic focus. Since most of stomach cancer cell strains established so far cause no peritoneal metastasis, this high peritoneal metastatic cell strain OCUM-2MD3 has been used as a typical model of peritoneal metastasis of stomach cancer.

Existence of several molecules, which may be associated with peritoneal metastasis, has been revealed using the high peritoneal metastatic cell strain OCUM-2MD3 as the experimental material. For example, the amount of "E-cadherin", a cell adhesion factor, has been found to be reduced in the OCUM-2MD3 strain compared with the parent strain OCUM-2M, verifying the fact that OCUM-2MD3 cells have a low cell adhesion capability so that they are apt to leave from the primary focus. Furthermore, production of MMP-1, one of the extracellular matrix degradation enzymes MMPs that are closely associated with cancer cell invasion, are elevated in the OCUM-2MD3 strain compared with those of the parent strain OCUM-2M. Since MMP-1 is an enzyme acting on type 1 and type 3 collagens which are characteristic constitutive proteins of gastric parietis, the elevated production of MMP-1 may prove the tendency of the OCUM-2MD3 strain to leave from the primary focus into the peritoneal cavity. In fact, the OCUM-2MD3 strain shows a higher invasive capability as compared with the parent strain OCUM-2M when their invasive capabilities into the matrigel are compared by

an invasion assay.

On the other hand, CD44H and a β_1-integrin family have been revealed as factors supporting the adhesion of cancer cells to the peritoneum. Expression levels of these adhesion molecules are elevated in OCUM-2MD3 cells. It has been suggested that hyaluronic acid existing in the peritoneal mesothelium, and fibronectin and laminin constituting the peritoneal interstitum function as the ligand of CD44 and β_1-integrin family, respectively, thereby assisting the adhesion of the OCUM-2MD3 cells to the peritoneum (Jap. J. Cancer Res. 87: 1235-1244, 1996; Br. J. Cancer 74: 1406-1412, 1996).

Although the existence of various factors related to peritoneal metastasis has been revealed, these findings are far to result in the treatment of metastasis. Therefore, elucidation of novel factors, which may lead to the treatment of peritoneal metastasis, is strongly desired.

Disclosure of the Invention

The object of the present invention is to provide genes which expression level varies depending on carcinogenesis of gastric tissues and malignancy of stomach cancer.

The present inventors speculated that it is possible to find genes which expression level is changes in cancer cells, by comparing the gene expression profile between stomach cancer cells and normal cells. To reveal a gene which expression level changes in stomach cancer among human genes, the number of which is predicted to be several tens of thousand to several hundreds of thousand, a technique enabling comparative analysis of the expression levels of many genes at the same time is needed. Comparison of the gene expression levels is usually performed by an analyzing method referred to as the differential analysis, for which the Northern blot and RT-PCR methods have been conventionally used. However, the application of such techniques to all the genes expressed in cells as the subject of analysis requires enormous amount of labor and time, which is unpractical. Besides these methods, the differential display method (DD method) has been also known in the field. However, the number of genes that can be identified by the DD method is far from enough, and it also requires advanced

technique and lots of labor.

DNA chips are constituted of microarrays comprising several tens of thousand to several hundreds of thousand oligonucleotides or polynucleotides, which nucleotide sequences are already known, immobilized at a high density. The target to be analyzed is fluorescence-labeled, and is brought into contact with these probe arrays. cDNAs derived from a variety of cells, and cRNAs synthesized using cDNAs as the template are generally used as a target. After hybridization, the microarray is thoroughly washed, and scanned for the fluorescence label remaining on the array to detect the probe to which the target has hybridized and to determine the amount of the hybridized target. These operations can be carried out simply in a short time. Furthermore, information on the presence and amounts of individual nucleotide sequences ranging from several tens of thousands to several hundreds of thousands kinds of sequences can be obtained by a single analysis. The information thus obtained is referred to as an expression profile. The differential analysis may be performed by comparing the expression profiles among different cells to select nucleotide sequence(s), which expression patterns are different among cells, using DNA chips.

Detection of changes in expression levels of a gene specifically found in stomach cancer cells can be carried out, for example, by comparing the expression level of the gene in a stomach cancer cell and normal cell, as well as in a primary stomach cancer cell and metastatic cancer cell and such as a combination to identify the gene of which expression level changes specifically in stomach cancer cells or by malignant alteration of the stomach cancer. Based on such viewpoint, the inventors compared cancerous tissues excised from cancer patients with a normal tissue portion derived from the same kind of tissue as the cancerous tissue and with a metastatic cancer tissue.

Alternatively, isolation of a gene specifically expressed in the OCUM-2MD3 cell strain, the high peritoneal metastatic type cell strain, may enable to elucidate a factor associated with the peritoneal metastasis in scirrhus gastric cancer. The present inventors presumed it might be possible to efficiently isolate a target gene by comparing

the OCUM-2MD3 strain with its parent OCUM-2M strain that is different from the former strain only in the capability to cause peritoneal metastasis and which commonly shares basic genetic characteristics.

By screening a cDNA library based on the selected nucleotide sequences, it is possible to finally isolate genes which expression level is specifically altered in cancer cells. The cDNA library can be synthesized from cancer cells or normal cells by well-known methods. However, cloning and sequencing of genes using a cDNA library synthesized by ordinary methods are a time-consuming work comprising repeated steps of sequencings of a plurality of positive clones and assembling of these clones. The present applicants have found out that the screening can be very quickly performed by using a full-length cDNA library constructed by the present applicants as the cDNA library and a database comprising the nucleotide sequences of the cDNA library.

The full-length cDNA library used in the present invention was synthesized by the oligo-capping method [K. Maruyama and S. Sugano, Gene, 138: 171-174 (1994); Y. Suzuki et al., Gene, 200: 149-156 (1997)], and has a high fullness ratio. All of the 5′-end nucleotide sequence and most of the 3′-end nucleotide sequence of the library are already determined. Additionally, the determination of the entire nucleotide sequence are also in progress. Furthermore, the results of homology searches between these elucidated partial nucleotide sequences or the full-length nucleotide sequence with known nucleotide sequences of genes and ESTs are also included in the database.

It is possible to obtain a full-length cDNA clone by detecting a clone comprising a nucleotide sequence identical to the nucleotide sequence selected based on the results of differential analysis with the DNA chips using this database without cloning by hybridization. The present invention has been completed under these circumstances. More specifically, the present invention relates to the following polynucleotides and proteins encoded by these polynucleotides, as well as uses thereof.

Table 1: SEQ ID NOs of nucleotide sequences and corresponding amino acid sequences of the present invention

Sequence name	Nucleotide sequence	Amino acid sequence
C-HEMBA1002150	1	2
C-HEMBA1002417	3	4
C-HEMBA1002475	5	6
C-HEMBA1002716	7	
C-HEMBA1003615	8	9
C-HEMBA1003805	10	11
C-HEMBA1004055	12	13
C-HEMBA1004669	14	15
C-HEMBA1004889	16	17
C-HEMBA1005621	18	19
C-HEMBA1006676	20	21
C-HEMBA1007085	22	23
C-HEMBB1001294	24	25
C-HEMBB1001482	26	27
C-HEMBB1002600	28	29
C-MAMMA1000284	30	31
C-MAMMA1000416	32	33
C-MAMMA1001388	34	35
C-MAMMA1002143	36	37
C-MAMMA1002351	38	39
C-MAMMA1002461	40	41
C-NT2RM1000039	42	43
C-NT2RM1000055	44	45
C-NT2RM1000355	46	47
C-NT2RM1001105	48	49
C-NT2RM2000101	50	51
C-NT2RM2000522	52	53
C-NT2RM2001345	54	55
C-NT2RM2001637	56	57
C-NT2RM2001696	58	59
C-NT2RM4000027	60	61
C-NT2RM4000514	62	63
C-NT2RM4001155	64	65
C-NT2RM4001382	66	67
C-NT2RM4002390	68	69
C-NT2RM4002593	70	
C-NT2RP2000289	71	72
C-NT2RP2000459	73	74
C-NT2RP2001327	75	76
C-NT2RP2001420	77	78
C-NT2RP2002193	79	80
C-NT2RP2002208	81	82
C-NT2RP2002606	83	84
C-NT2RP2003272	85	86
C-NT2RP2004013	87	88
C-NT2RP2004242	89	90
C-NT2RP2005360	91	92
C-NT2RP3000109	93	94

C-NT2RP3000605	95	96
C-NT2RP3001730	97	98
C-NT2RP3002273	99	100
C-NT2RP3002399	101	102
C-NT2RP3002818	103	104
C-NT2RP3002948	105	106
C-NT2RP3003290	107	108
C-NT2RP3003876	109	110
C-NT2RP3004041	111	112
C-NT2RP4000973	113	114
C-OVARC1000781	115	116
C-OVARC1001270	117	118
C-OVARC1001726	119	120
C-PLACE1000133	121	122
C-PLACE1000786	123	124
C-PLACE1001845	125	126
C-PLACE1004506	127	128
C-PLACE1005409	129	
C-PLACE1005603	130	131
C-PLACE1006037	132	133
C-PLACE1006469	134	135
C-PLACE1008947	136	137
C-PLACE3000242	138	139
C-PLACE4000052	140	141
C-THYRO1000401	142	143
C-Y79AA1000258	144	145
C-Y79AA1000784	146	147
C-Y79AA1001781	148	149

(1) A polynucleotide selected from the group of:

(a) a polynucleotide comprising any of the nucleotide sequences of SEQ ID NOs shown in Table 1;

(b) a polynucleotide encoding a protein comprising any of the amino acid sequences of SEQ ID NOs shown in Table 1;

(c) a polynucleotide encoding a protein comprising any of the amino acid sequences of SEQ ID NOs shown in Table 1 in which one or more amino acids are substituted, deleted, inserted and/or added, and is functionally equivalent to a protein consisting of said amino acid sequences; and,

(d) a polynucleotide hybridizing under a stringent condition to a polynucleotide consisting of any of the nucleotide sequences of SEQ ID NOs shown in Table 1, which encodes a protein functionally equivalent to the protein consisting of the amino acid sequence encoded by said nucleotide sequence;

(2) a polynucleotide encoding a partial peptide of the protein

encoded by the polynucleotide of (1);

(3) a protein or a partial peptide thereof, which is encoded by the polynucleotide of (1) or (2);

(4) a vector containing the polynucleotide of (1) or (2);

(5) a transformant harboring the polypeptide of (1) or (2), or the vector of (4);

(6) a method for preparing the protein or partial peptide of (3), which comprises the steps of:

(a) culturing the transformant of (5); and,

(b) recovering the expressed product;

(7) a polynucleotide comprising at least 15 nucleotides complementary to the polynucleotide of (1) or (2), or a complementary strand thereof;

(8) an antibody against the protein or partial peptide of (3);

(9) an immunological assay method comprising the step of observing an immunological reaction of the protein of (3) with the antibody of (8);

(10) a method of screening for a compound that controls the expression of the polynucleotide of (1), which comprises the steps of:

(a) contacting a candidate compound with stomach cancer cells;

(b) comparing the expression level of a gene, which comprises the nucleotide sequence according to a SEQ ID NO shown in Table 1, in the stomach cancer cells with that in control cells; and,

(c) selecting the candidate compound which alters the expression level of the gene;

(11) an use of a compound for controlling carcinogenesis and metastasis of stomach cancer, wherein the compound can be obtained by the method of (10);

(12) a method for detecting stomach cancer comprising the steps of:

(a) measuring the polynucleotide of (1) in specimens from living bodies; and,

(b) correlating the measured results of (a) with the presence of stomach cancer;

(13) a method for detecting stomach cancer comprising the steps

of:

(a) measuring proteins and/or partial polypeptides of (3) in specimens from living bodies; and,

(b) correlating the measured results of (a) with the presence of stomach cancer.

The present invention relates to isolated polynucleotides related with stomach cancer. A polynucleotide provided by the present invention comprises the nucleotide sequence of a gene, which expression level is specifically altered in stomach cancer compared with that in normal tissue, and/or in the metastatic cancer compared with that in primary cancerous tissue. Alternatively, a polynucleotide provided by the present invention comprises the nucleotide sequence of a gene, which expression level is specifically altered in a stomach cancer cell strain, which is apt to cause peritoneal metastasis.

The polynucleotide of the present invention includes, besides DNA and cDNAs, genomic DNAs, chemically synthesized DNAs, and RNAs. Furthermore, the polynucleotide of the present invention may include not only naturally occurring nucleotides but also artificially synthesized nucleotide derivatives and labeled nucleotides. The term "oligonucleotide" refers to polynucleotides herein. The term "oligonucleotide" refers to polynucleotides with a relatively short nucleotide chain. The term "polynucleotide" encompasses the oligonucleotides. Polynucleotides of the present invention include, for example, vectors, autonomously replicating plasmids and viruses, recombinant polynucleotides incorporated in genomic DNAs of prokaryotes or eukaryotes, and recombinant polynucleotides existing as independent molecules from other sequences. Furthermore, the polynucleotide of the present invention also includes recombinant DNAs existing as a portion of a hybrid gene encoding an additional polypeptide sequence.

SEQ ID NOs of desirable nucleotide sequences of polynucleotides provided by the present invention are as shown in Table 1, in which SEQ ID NOs of amino acid sequences of proteins encoded by these nucleotide sequences are also shown. The present invention provides proteins comprising these amino acid sequences.

Expression profiles of genes shown in Table 1 are represented

in Table 2. Genes described in the column of the method for selection in Table 2 as "5a" (5 times or more in #5 than in #3), "5b" (5 times or more in #5 than in #12) or "5c" (3 times or more than in #3, and 3 times or more than in #12 in #5) indicate that the expression level of the gene in human stomach cancer focus formed after the subcutaneous transplantation of the human stomach cancer cells (#5) to SCID mice has increased 5 times or more compared with the expression level thereof in a normal gastric mucosa (#3 or #12) or 3 times or more compared with that in both of the normal gastric mucosa, #3 and #12, and thus, were selected as genes, which expression levels increased in stomach cancer. Genes meeting these requirements include the following: MAMMA1002351, NT2RP2001327, NT2RM1000355, Y79AA1000784, NT2RM4001382, NT2RM1000055, PLACE1008947, MAMMA1002461, NT2RP3004041, NT2RM2001637, PLACE1006469, HEMBA1002417, HEMBB1002600, NT2RM4002390, Y79AA1000258, NT2RM4000027, MAMMA1002143, NT2RP4000973, NT2RP2005360, HEMBA1003615, NT2RM2000522, HEMBA1002475, NT2RP2004242, NT2RM2001637, Y79AA1000784, NT2RM4001382, HEMBA1004889, HEMBA1006676, NT2RM2001696, NT2RM4002593, Y79AA1001781, HEMBA1003805, NT2RP2002606, NT2RP3003876, OVARC1001726, HEMBA1005621, NT2RM4000514, NT2RM1000039, MAMMA1001388, MAMMA1001388, HEMBA1007085, NT2RM2001345, NT2RP2000289, NT2RM4001155, and NT2RP3002818.

Furthermore, genes described in the column of the method for selection in Table 2 as "13a" (5 times or more than #3 in #13), "13b" (5 times or more in #13 than in #12) or "13c" (3 times or more than in #3, and 3 times or more than in #12 in #13), "18a" (5 times or more in #18 than in #3), "18b" (5 times or more in #18 than in #12) or "18c" (3 times or more than in #3 and 3 times or more than in #12 in #18) indicate that the expression levels of the genes in clinical specimens, #13 or #18, derived from stomach cancers increased 5 times or more compared with the expression level thereof in the normal gastric mucosa, #3 or #12, or 3 times or more compared with those in both of the normal gastric mucosa, #3 and #12, and thus, were selected as genes, which expression levels increased in stomach cancer. Genes meeting these requirements include the following: HEMBB1001294,

NT2RP2001327, NT2RP2000459, Y79AA1000784, NT2RM4001382, HEMBA1002716, NT2RP2002193, THYRO1000401, OVARC1000781, PLACE4000052, NT2RP3002948, PLACE1001845, PLACE1006469, PLACE1000786, MAMMA1000416, PLACE1005409, NT2RP3000605, NT2RM4002390, HEMBA1004055, PLACE1005603, HEMBA1002150, Y79AA1000258, NT2RM1001105, PLACE1006037, OVARC1001270, HEMBB1001482, MAMMA1000416, PLACE1000133, NT2RP2004013, PLACE3000242, NT2RP3003290, HEMBA1006676, NT2RM2001696, HEMBA1007085, NT2RP3000109, PLACE1004506, PLACE1005409, NT2RP2003272, HEMBA1005621, NT2RP3002399, NT2RM2000101, NT2RP2002208, NT2RM4000514, NT2RP3002273, MAMMA1000284, HEMBA1007085, HEMBA1004669, and NT2RP3001730.

Furthermore, genes described in the column of the method for selection in Table 2 as "14" indicate that the expression levels thereof in the metastatic focus in the lymph node (#14) have been increased 5 times or more than that in the stomach cancer tissue (#13), and thus, were selected as genes, which expression levels increased in stomach cancer. Genes meeting these requirements include the following: NT2RP2001420, PLACE1000786 and MAMMA1002143.

In addition, gene "MAMMA1001388" having the nucleotide sequence set forth in SEQ ID NO: 34 (amino acid sequence: SEQ ID NO: 35) has been found that the expression level increased 5 times or more in the stomach cancer cell strain OCUM-2MD3 (D3), a strain with higher peritoneal metastatic capability than the parent stomach cancer cell strain OCUM-2M (2M), and thus, was selected as a gene, which expression level increases in stomach cancer.

There is no particular limitation on the type of polynucleotides of the present invention, so long as they can encode proteins of the present invention, which includes, besides cDNAs, genomic DNAs, chemically synthesized DNAs, etc. Polynucleotides of the present invention also include those having any nucleotide sequences based on the degeneracy of genetic codes so long as these polypeptides can encode the protein of the present invention. Polynucleotides encoding proteins of the present invention can be isolated by a standard method, such as a hybridization method using the polynucleotide sequences according to SEQ ID NOs shown in Table 1 or portions thereof as the

probe, and the PCR method using primers designed based on the information of these polynucleotide sequences.

Genes comprising the nucleotide sequences according to SEQ ID NOs shown in Table 1 include those found in the stomach cancer cells of high malignancy causing metastasis to lymph nodes and peritoneal metastasis. Therefore, expression analysis of these genes may reveal the malignancy of the cancer cells, thereby providing critical information for planning treatment strategies.

Peritoneal metastasis of stomach cancer is thought to be established by the steps of: (1) the primary focus inside the gastric parietal tissue proliferates and invades to the outside of the gastric parietal tissue, and the focus further separates from the serosa and floats in the peritoneal cavity; and (2) the floating cells implant to the peritoneum to proliferate therein. The genes of the present invention were isolated from a high peritoneal metastatic cell strain, and thus, are considered to be important genes supporting this series of processes. Therefore, inhibition of functions of these genes may enable prophylaxis and treatment of peritoneal metastasis. In addition, the genes of the present invention specific to the high peritoneal metastatic cell strain and the proteins encoded by these genes are useful as indicators to assess the malignancy of stomach cancer. Herein, malignancy of stomach cancer means the capability of the cancer to cause peritoneal metastasis and metastases to the lymph nodes.

Furthermore, the genes of the present invention may be also used for prophylaxis and treatment of peritoneal metastasis or prediction of malignancy similarly in cancers of digestive organs such as pancreatic cancer besides stomach cancer. Since metastases to peritoneum and lymph nodes are the malignant alteration step commonly observed among various cancers of digestive organs, the genes of the present invention may play a similar role in other solid tumors.

For example, it was revealed that the expression level of gene "MAMMA1000416" comprising a nucleotide sequence according to SEQ ID NO: 32 (amino acid sequence: SEQ ID NO: 33) is significantly elevated not only in stomach cancer but also in hepatoma, indicating a possibility that the expression level of the gene of the present invention is

also elevated in solid tumors other than stomach cancer.

As described above, it can be said that the genes comprising the nucleotide sequence provided by the present invention are closely associated with carcinogenesis and malignancy of stomach cancer. Therefore, it is considered that diagnosis and treatment of stomach cancer is possible by controlling the expression of these genes and the actions of the proteins encoded by these genes. More specifically, the present invention relates to compounds capable of controlling the expression of the genes of the present invention and a method of screening for such compounds.

More specifically, the progress and metastasis of stomach cancer can be effectively suppressed by inhibiting the expression of a gene of the present invention *in vivo*. Alternatively, the suppression of stomach cancer can be also achieved by inhibiting the action of a protein of the present invention. Suppression of the expression of above-mentioned genes can be achieved by inhibiting the expression using antisense nucleic acid drugs or decoy nucleic acids after determining the transcriptional control region thereof. Effective inhibition of the action itself of the protein is achieved by administering compounds binding to the protein so as to modify the three-dimensional structure of the active site of the protein, or by interfering the binding of the protein to its target compound.

Furthermore, cancer vaccine may be developed utilizing the proteins of the present invention. More specifically, immunological exclusion mechanisms against stomach cancer will be strengthened by inducing an immune response to a protein encoded by a gene of the present invention or fragments thereof. Such an immune response can be caused by administering the protein of the present invention or fragments thereof into the living body. Administration of proteins into the living body can be achieved by a direct administration of the protein, or by the transduction and expression of a gene encoding the protein. The desired gene can be transduced using adenoviral vectors and retroviral vectors according to conventional methods.

The protein encoded by the polynucleotide of the invention can be prepared as a recombinant protein or as a natural protein. For example, the recombinant protein can be prepared by inserting the

polynucleotide encoding the protein of the invention into a vector, introducing the vector into an appropriate host cell and purifying the protein expressed within the transformed host cell, as described below. Alternatively, *in vitro* translation (See, for example, "On the fidelity of mRNA translation in the nuclease-treated rabbit reticulocyte lysate system." Dasso M.C., and Jackson R.J. (1989) Nucleic Acids Res. 17: 3129-3144) may be used for preparing the protein of the invention. In contrast, the natural protein can be prepared, for example, by utilizing an affinity column to which an antibody against the protein of the invention (Current Protocols in Molecular Biology (1987) Ausubel et al. edit, John Wiley & Sons, Section 16.1-16.19) is attached. The antibody used for affinity purification may be either a polyclonal antibody, or a monoclonal antibody.

The present invention also includes polynucleotides encoding not only proteins comprising the amino acid sequences according to SEQ ID NOs shown in Table 1, but also proteins functionally equivalent to these proteins. Herein, the term "functionally equivalent" indicates that the subject protein causes carcinogenesis or malignant alteration of stomach cancer, and in such a case, the subject protein is determined to be functionally equivalent to the protein of the present invention.

In the present invention, the capability of a gene to cause carcinogenesis can be confirmed by detecting carcinogenesis of a host cell by the transformation of the gene. On the other hand, the capability of a gene to cause malignant alteration can be identified by utilizing the fact that a cancer cell strains with no metastatic ability acquires a metastatic ability by the transformation of the gene as an indicator. For example, a cell strain with a low or no metastatic capability, such as the stomach cancer cell strain OCUM-2M, can be used for the detection of malignant alteration due to the transformation of the gene.

Proteins functionally equivalent to the proteins of the present invention can be prepared by those skilled in the art, for example, by using a method for introducing mutations into an amino acid sequence of a protein (for example, site-directed mutagenesis (Current Protocols in Molecular Biology, edit, Ausubel et al., (1987) John

Wiley & Sons, Section 8.1-8.5). Besides, such proteins can be generated by spontaneous mutations in nature. The present invention comprises the proteins having one or more amino acids substitutions, deletions, insertions and/or additions in the amino acid sequences of the proteins of the present invention (SEQ ID NOs in Table 1), as far as the proteins have an equivalent functions to those of the proteins identified in the present Examples described below.

There is no limitation in numbers and sites of mutations of amino acid residues in proteins so long as they retain their own functions. The number of mutations is typically less than 10%, preferably less than 5%, and more preferably less than 1% of the total amino acid residues. In view of retaining the function of the protein, the amino acid residue to be substituted is preferably substituted with an amino acid residue having a similar property as that of the amino acid residue to be substituted. For example, Ala, Val, Leu, Ile, Pro, Met, Phe and Trp are assumed to have similar properties to one another because they are all classified into a group of non-polar amino acids. Similarly, substitution can be performed among non-charged amino acid such as Gly, Ser, Thr, Cys, Tyr, Asn, and Gln, acidic amino acids such as Asp and Glu, and basic amino acids such as Lys, Arg, and His.

In addition, proteins functionally equivalent to the proteins of the present invention can be isolated by using techniques of hybridization or gene amplification known to one skilled in the art. Specifically, using the hybridization technique (Current Protocols in Molecular Biology edit. Ausubel et al., (1987) John Wiley & Sons, Section 6.3-6.4)), one skilled in the art can usually isolate a polynucleotide highly homologous to the polynucleotide encoding the protein identified in the present Example based on the identified polynucleotide sequence (Table 1) or a portion thereof and obtain the functionally equivalent protein from the isolated polynucleotide. The present invention includes proteins encoded by the DNAs hybridizing with the polynucleotide encoding the proteins identified in the present Example, as far as the proteins are functionally equivalent to the proteins identified in the present Example. Organisms from which the functionally equivalent proteins are isolated are illustrated by vertebrates such as human, mouse, rat, rabbit, pig and bovine, but

are not limited thereto. Such genes retain a high homology at the nucleotide sequence level.

Stringent conditions of hybridization for isolating polynucleotides encoding functionally equivalent proteins include a washing condition usually "1x SSC, 0.1% SDS at 37°C" or thereabout. A more stringent washing condition is "0.5x SSC, 0.1% SDS at 42°C" or thereabout, and a further more stringent washing condition is "0.1x SSC, 0.1% SDS at 65°C" or thereabout. The more stringent the conditions of the hybridization are the higher the homology of the isolated polynucleotide to the probe sequence will be. However, the above-mentioned combinations of conditions concerning SSC, SDS, and temperature are merely examples, and one skilled in the art can accomplish a similar stringency as described above by appropriately combining the aforementioned elements and others that determine the stringency of hybridization (for example, the concentration of the probe, length of the probe, reaction time of the hybridization, etc.).

Proteins isolated using such hybridization techniques usually have a high homology at the amino acid sequences level compared with the protein of the present invention according to SEQ ID NOs shown in Table 1. Herein, high homology means that the sequence is at least 60% or more, preferably 70% or more, more preferably 80% or more (for example, 90% or more) identical to the other sequence. The "percent identity" of two amino acid sequences or of two nucleic acids is determined using the algorithm of Karlin and Altschul (Proc. Natl. Acad. Sci. USA 90:5873-5877, 1993). Such an algorithm is incorporated into the BLASTN and BLASTX programs of Altschul et al. (J. Mol. Biol. 215: 403-410, 1990). BLAST nucleotide searches are performed with the BLASTN program, score = 100, wordlength = 12. BLAST protein searches are performed with the BLASTX program, score = 50, wordlength = 3. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs are used. Specific methods of these analyzing methods are well known (See http://www.ncbi.nlm.nih.gov).

It is also possible to obtain a protein functionally equivalent to the protein encoded by a gene identified in the Example of the present invention using the gene amplification technique (PCR) (Current Protocols in Molecular Biology edit. Ausubel et al. (1987)

Publish. John Wiley & Sons Section 6.1-6.4) comprising the steps of: (1) designing a primer based on a portion of the nucleotide sequence (Table 1) identified in the Example; (2) isolating polynucleotide fragments comprising a nucleotide sequence with a high homology to the nucleotide sequences or portions thereof; and (3) obtaining the protein functionally equivalent to the protein encoded by a gene identified in the Examples based on the polynucleotide fragments.

It is also possible to isolate a polynucleotide encoding a functionally equivalent protein by a homology search on a computer in addition to performing hybridization and PCR as described above. Polynucleotides encoding proteins of the present invention may be homologous genes of the genes comprising the nucleotide sequences according to Table 1 preserved among species, or genes that are not homologous to these sequences but that are similar to them, and which are highly homologous to the polynucleotides encoding the proteins of the present invention according to SEQ ID NOs shown in Table 1.

The present invention also provides partial peptides of the proteins of the present invention. The partial peptides are useful as the antigen to obtain antibodies to the proteins of the present invention. In particular, a partial peptide having a low homology to other proteins, and comprising a unique amino acid sequence of the protein of the present invention is expected as an immunogen to produce antibodies with a high specificity to the protein of the present invention.

The partial peptides of the present invention comprise at least 7 amino acid residues, preferably 9 or more amino acid residues, more preferably 12 or more amino acid residues, and further more preferably 15 or more amino acid residues. The partial peptides of the present invention can be produced, for example, by genetic engineering techniques, conventional peptide synthesis methods, and by digesting the protein of the present invention with an appropriate peptidase.

The present invention also provides expression vectors containing any of the above-mentioned polynucleotides. The vector of the invention is not limited as long as it contains the inserted polynucleotide stably. For example, if *E. coli* is used as a host, vectors such as pBluescript vector (Stratagene) are preferable as

the cloning vector. To produce the protein of the invention, expression vectors are especially useful. Any expression vector can be used as far as it is capable of expressing the protein *in vitro*, in *E. coli*, in cultured cells, or *in vivo*. For example, pBEST vector (Promega) is preferable for *in vitro* expression, pET vector (Novagen) for *E. coli*, pME18S-FL3 vector (GenBank Accession No. AB009864) for cultured cells, and pME18S vector (Mol. Cell. Biol. (1988) 8: 466-472) for *in vivo* expression. To insert the polynucleotide of the invention, ligation utilizing restriction sites can be performed according to the standard method (Current Protocols in Molecular Biology (1987) Ausubel et al. edit, John Wiley & Sons, Section 11.4-11.11).

Furthermore, the present invention relates to transformants harboring the above-mentioned polynucleotides or any one of the aforementioned expression vectors, and also to a method for preparing the protein of the present invention comprising the steps of: (1) culturing the transformant; and (2) isolating the protein of the present invention from the culture. There is no particular limitation on the type of host cells to which the vector of the present invention is transferred, and a variety of host cells can be used according to a predetermined purpose. Examples of eukaryotic cells for expression of the protein at large quantities are COS cells, CHO cells, etc.

Transduction of a vector to a host cell may be performed by, for example, the calcium phosphate precipitation method, the electropulse poration method (electroporation method) (Current Protocols in Molecular Biology edit. Ausubel et al. (1987) Publish. John Wiley & Sons Section 9.1-9.9), the lipofectamin method (GIBCO-BRL), the microinjection method, and so on. The present invention provides proteins or partial peptides thereof prepared by the above-mentioned methods.

Procedures such as DNA cloning, construction of each plasmid, transfection of hosts, culturing of transformants, and recovery of the proteins from the culture required for implementing the present invention can be performed according to conventional methods or methods described in the literature (Molecular Cloning, T. Maniatis, et al., CHS Laboratory (1983) DNA Cloning, DM. Glover, IRL PRESS (1985), and others).

The host cells of the present invention also include cells used for the purpose of functional analysis of the genes of the present invention, and those used for screening of inhibitors of a gene function using the genes. The transduction of a vector to a host cell may be performed by, for example, the calcium phosphate precipitation method, the electropulse poration method (electroporation method) (Current Protocols in Molecular Biology edit. Ausubel et al. (1987) Publish. John Wiley & Sons Section 9.1-9.9), the lipofectamin method (GIBCO-BRL), the microinjection method, etc. Preparation of the protein of the present invention from transformants can be carried out using methods for isolating and purifying proteins known to those skilled in the art.

The present invention also provides polynucleotides consisting of at least 15 nucleotides complementary to the polynucleotide comprising the nucleotide sequence according to SEQ ID NOs shown in Table 1 or complementary strands thereof. Herein, the term "complementary strand" is defined as one strand of a double-stranded polynucleotide composed of the base pairs, A:T and G:C, to the other strand. "Complementary" is also defined as not only those completely matching within a continuous region of at least 15 nucleotides, but also those having a homology of at least 70%, preferably at least 80%, more preferably 90%, and further more preferably 95% or more within that region. The homology can be determined using the algorithm described herein.

Such polynucleotides may be used as probes for detection and isolation of the DNAs and RNAs encoding the proteins of the present invention, and also as primers for amplifying the polynucleotides of the present invention. When used as primers, such oligonucleotides usually have a chain length of 15 bp to 100 bp, preferably 15 bp to 35 bp. On the other hand, when used as probes, such polynucleotides comprising at least a part or the entire sequence of the polynucleotide of the present invention and have a chain length of at least 15 bp. When used as primers, such polynucleotides are complementary at the 3'-end region, and restriction enzyme recognition sequences and tags may be added to the 5'-end.

The polynucleotides of the present invention can be used for

detecting the expression of the genes of the present invention or for quantifying the amount of the gene expression. For example, it is possible to examine the expression level of a gene by the Northern hybridization and RT-PCR using the polynucleotide of the present invention as the probe and primer, and additionally, aberration of a sequence can be examined and diagnosed by the RFLP analysis, the SSCP, sequencing, and so on, after amplifying DNAs of the present invention or the expression control region thereof by the polymerase chain reaction (PCR), genomic DNA-PCR or RT-PCR, using the polynucleotide of the present invention as the probe and primer.

Furthermore, antisense DNA suppressing the expression of a gene of the present invention are included in the "DNA having at least 15 nucleotides and which are complementary to the polynucleotide comprising a nucleotide sequence according to SEQ ID NOs shown in Table 1 or a complementary strand thereof". To effect as an antisense, the antisense DNA has a chain length of at least 15 bp or more, preferably 100 bp or more, more preferably 500 bp or more, and usually a chain length less than 3000 bp, more preferably less than 2000 bp.

Such antisense DNAs can be applied to gene therapy of progressive and metastatic stomach cancer. The antisense DNAs may be prepared based on the DNA sequence information according to SEQ ID NOs shown in Table 1 by the phosphorothioate method (Stein, 1988 Physichochemical properties of phosphorothioate oligodeoxynucleotides. Nucleic Acids Res 16, 3209-21 (1988)), etc.

When using the polynucleotides or antisense DNAs of the present invention in gene therapy, they are administered to patients by the *ex vivo* and *in vivo* methods utilizing viral vectors such as retroviral vector, adenoviral vector, adeno-associated viral vector, and such, and non-viral vector such as liposomes.

The present invention also provides antibodies binding to the proteins of the present invention. There is no particular limitation on the type of antibody of the present invention, and includes polyclonal and monoclonal antibodies or also portions thereof having the antigen-binding capability, and furthermore, antibodies of all classes. The antibodies of the present invention also include a specific antibody such as humanized antibodies.

The polyclonal antibody of the invention can be obtained according to the standard method by synthesizing an oligopeptide corresponding to the amino acid sequence and immunizing rabbits with the peptide (Current Protocols in Molecular Biology (1987) Ausubel et al. edit, John Wily & Sons, Section 11.12-11.13). On the other hand, the monoclonal antibody of the invention can be obtained according to the standard method by purifying the protein expressed in *E. coli*, immunizing mice with the protein, and producing a hybridoma cell by fusing the spleen cells and myeloma cells (Current Protocols in Molecular Biology (1987) Ausubel et al. edit, John Wily & Sons, Section 11.4-11.11).

The antibody binding to the protein of the present invention can be used for purification of the protein of the invention, and also for detection and/or diagnosis of the abnormalities of the expression and structure of the protein. More specifically, the identification of cancer and examination/diagnosis of malignancy thereof can be performed, for example, by extracting proteins from tissues, blood or cells, and through following detection of the proteins of the present invention by the Western blotting, immunoprecipitation, the ELISA methods, etc.

For example, the existence of the polynucleotides and proteins of the present invention or fragments thereof in a tissue specimen indicates that the tissue is derived from stomach cancer. Furthermore, the presence of the polynucleotides and proteins of the present invention or fragments thereof in the blood can be also used as an indicator of stomach cancer. All of the polynucleotides of the present invention comprise nucleotide sequences of genes of which expression were confirmed to increase in stomach cancer cells. Therefore, in case where the measured values of the polynucleotides and proteins of the present invention or fragments thereof are increased compared with that of normal individuals, the presence of stomach cancer is suspected. mRNA can be mentioned as an example of the polynucleotide of the present invention enabling detection of stomach cancer. mRNA can be used as an indictor of stomach cancer by detecting the presence in blood and cells by techniques such as RT-PCR. Alternatively, the protein of the present invention or fragments thereof detected by

immunological techniques known in the field may be also used as an indicator of stomach cancer.

Antibodies binding to the proteins of the present invention may be used for the purpose of stomach cancer treatment, and so on. The proteins encoded by the genes of the present invention are highly expressed in stomach cancer and stomach cancer with high malignancy. Thus, antibodies recognizing these proteins are useful for immunological treatments of stomach cancer. Furthermore, a missile therapy of stomach cancer may be realized by binding an anticancer agent to the antibodies targeting these proteins. When the antibodies are used for the treatment of human patients, human antibodies or humanized antibodies are preferred to reduce immunogenicity. The human antibodies can be prepared by immunizing a mouse whose immune system is replaced with that of human ("Functional transplant of megabase human immunoglobulin loci recapitulates human antibody response in mice" Mendez M.J. et al. (1997) Nat. Genet. 15: 146-156). The humanized antibodies can be prepared by recombination of the hypervariable region of a monoclonal antibody (Methods in Enzymology (1991) 203: 99-121).

Further, the present invention provides a method of screening for compounds to control the activity of the protein of the present invention. Since the gene of the present invention is associated with carcinogenesis and malignancy of stomach cancer, compounds that suppress the activity of the product of the gene are useful as therapeutic agents to suppress stomach cancer and metastasis thereof. This screening method comprises the steps of:

(a) contacting a candidate compound with stomach cancer cells;
(b) comparing the expression level of a gene comprising a nucleotide sequence according to a SEQ ID NO shown in Table 1; and,
(c) selecting the candidate compound that reduces the expression level of the gene.

Stomach cancer tissues excised from patients, and stomach cancer cell strains can be used as the stomach cancer cell specimen in the screening method of the present invention. Alternatively, cells to which a gene of the present invention has been artificially transduced can be also used as materials for the screening. The expression level

of a gene comprising a nucleotide sequence according to a SEQ ID NO shown in Table 1 is used as an indicator in the screening method of this invention. Since the genes of the present invention are associated with carcinogenesis and metastasis of stomach cancer, the cell type and genes used as the indicator can be appropriately selected according to the purpose of the screening. For example, when the purpose is to control carcinogenesis, a gene of which high expression could be observed is selected as the indicator. Alternatively, when screening for compounds capable of controlling metastasis, a gene associated with malignancy is used as the indicator. The expression levels of the genes can be detected or quantified based on conventional methods such as the Northern blotting method, and RT-PCR method.

Samples to screen include, for example, cell extracts, expressed products from a gene library, synthesized low molecular compounds, synthesized peptides, and natural compounds, but are not limited thereto. A compound that is isolated by the above screening using a binding activity of the protein of the invention can also be used as a sample.

Compounds isolated by this screening method serves as a candidate inhibitor for the expression of the genes of the present invention. These compounds may be applied for the prophylaxis and treatment of stomach cancer and metastasis thereof with which the genes of the present invention are associated.

Compounds isolated by the screening method of the present invention can be administered to patients as it is, or after formulated into a pharmaceutical composition according to known methods. For example, a pharmaceutically acceptable carrier or vehicle, specifically sterilized water, saline, plant oil, emulsifier, or suspending agent can be mixed with the compounds appropriately. The pharmaceutical compositions can be administered to patients by a method known to those skilled in the art, such as intraarterial, intravenous, or subcutaneous injections. The dosage may vary depending on the weight or age of a patient, or the method of administration, but those skilled in the art can choose an appropriate dosage properly. If the compound is encoded by DNA, the DNA can be cloned into a vector for gene therapy, and used for gene therapy. The dosage of the DNA and the method of

its administration may vary depending on the weight or age of a patient, or the symptoms, but those skilled in the art can choose properly.

The invention is illustrated more specifically with reference to the following examples, but is not to be construed as being limited thereto.

Best Mode for Carrying out the Invention

[Example 1] Comparison of expression levels by differential analysis

Gene expression levels of following cells were analyzed. The results of normal and cancerous tissues, and those of cancerous tissue and metastatic lesion were compared to each other to select probe(s) hybridizing with genes of which expression level are altered 5-fold (or 3-fold) or more to each other. Numbers in the parentheses represent the numbers of the specimen.

Stomach cancer:

stomach cancer tissues: 2 cases (#13 and #18),

lymph node metastatic tissue derived from the same patient of the stomach cancer tissue #13: 1 case (#14),

normal gastric mucosa derived from the same patient of the stomach cancer tissue #13: 1 case: (#12),

stomach cancer cell strain OCUM-2M: 1 case,

stomach cancer cell strain, OCUM-2MD3, having a high peritoneal metastatic capability: 1 case,

stomach cancer transplanted to nude mice: 2 cases (#5 and #6), and

operation specimen of normal gastric mucosa: 1 case (#3).

The stomach cancer cell strain, OCUM-2M, established at the Department of Surgery 1, Osaka Prefecture University, and its substrain, OCUM-2MD3, which causes peritoneal metastasis at a high frequency (Br. J. Cancer 72: 1200-1210, 1995) were used as the cell strains. The following extraction and labeling of RNA as well as hybridization with microarrays were principally performed according to the

instructions of Affymetrix, Inc.

Poly(A)$^+$RNA was prepared from clinical specimens or a cell strain cultured in D-MEM medium containing 10% fetal bovine serum by the oligo (dT) cellulose spin column method (QuickPrep mRNA Purification kit, Pharmacia). Single-stranded cDNA was synthesized with the reverse transcriptase (Superscript RT II, BRL) using 1μg poly(A)$^+$RNA (1 μg) as a template and T7-incorporated oligo (dT) 24 as a primer, and then, double-stranded cDNA was further synthesized using the *E. coli* DNA ligase and *E. coli* DNA polymerase. The synthesized cDNA was extracted with phenol-chloroform according to a conventional method. cRNA was synthesized using this double-stranded cDNA as a template with the T7 RNA polymerase by the MEGAscript T7 kit (Ambion), wherein the cRNA was labeled by adding Biotin-11-CTP and Biotin-16-UTP as the labeling nucleotides. The synthesized cRNA was recovered using the RNeasy Mini Kit (QUIAGEN), and was purified by the SPIN-100 Columns (CLONTECH). The purified cRNA was heat-fragmented, and used for the hybridization with a cDNA oligonucleotide array (Affymetrix). The fragmentation of the cRNA was performed by adding the following fragmentation buffer (8 μl) to an RNase-free purified water (32 μl) containing the cRNA (20 μg) (final concentration of the cRNA: 0.5 μg/μl), followed by a treatment at 94°C for 35 min. The cRNA was fragmented into fragments of approximately 35 to 200 bp in size by the heat treatment.

5x fragmentation buffer

4.0 ml of 1 M Tris-acetate buffer (pH 8.1),

0.64 g of magnesium acetate, and

0.98 g of calcium acetate

was adjusted to 20 ml with the DEPC-treated H_2O.

The fragmented cRNA specimen was mixed as hybridization cocktail of the composition described below, treated once at 99°C for 5 min, and then, was placed on a heat block of 45°C for 5 min. An aliquot (200 μl) of the hybridization cocktail was added to each microarray, and allowed to hybridize at 45°C for 16 hrs. Five microarrays were

used for the hybridization: (1) HuGeneFL (formerly called Hu6800), about 6500 kinds of oligonucleotides derived from genes or ESTs are synthesized on the array; (2)-(5) Hu35Ks A, B, C and D, about 35,000 kinds (additional amount on A-D) of oligonucleotides derived from genes or ESTs are synthesized on these arrays in total. Herein, the GeneChip Fluidics Station 400 (Affymetrix) was used for all the steps of washings to fluorescence staining after the hybridization.

Hybridization cocktail:

fragmented cRNA 15 µg,

control oligonucleotide B2 (5 nM) 3 µl,

100x control cRNA cocktails 3 µl each,

salmon sperm DNA (10 mg/ml) 3 µl,

acetylated BSA (50 mg/ml) 3 µl, and

2x MES hybridization buffer 150 µl

was adjusted to a total volume of 300 µl.

After the completion of the hybridization, the hybridization cocktail was removed from the microarray, and a washing solution (250 µl) was added thereto. Following the removal of non-specific signals by washing, the streptavidin phycoerythrin (SAPE) was bound to the microarray. Furthermore, the fluorescence intensity was enhanced by antibodies to avidin and the streptavidin phycoerythrin again. Composition of the reaction solution used for the washing and fluorescence staining was as follows:

Washing solution:

12x MES stock buffer solution 83.3 ml,

5 M NaCl 5.2 ml,

10% Tween 20 1.0 ml, and

H_2O 910.5 ml.

Reaction solution for fluorescence staining:

2x staining buffer 300 μl,

H_2O 270 μl,

acetylated BSA (50 mg/ml) 24 μl, and

streptavidin phycoerythrin (1 mg/ml) 6 μl.

Anti-streptavidin antibody to enhance fluorescence (in 600 μl):

2x staining buffer 300 μl,

acetylated BSA (50 mg/ml) 24 μl,

normal goat IgG (10 mg/ml) 6.0 μl,

biotinylated antibody (0.5 mg/ml) 3.6 μl, and

H_2O 266.4 μl.

Streptavidin phycoerythrin to enhance fluorescence (in 1200 μl):

2 x staining buffer 600 μl,

acetylated BSA (50 mg/ml) 48 μl,

streptavidin phycoerythrin (1 mg/ml) 12 μl, and

H_2O 540 μl.

The fluorescence intensity of each fluorescence-stained microarray was measured with a confocal laser apparatus (HP Genearray scanner). Genes and ESTs on the five microarrays were compared for the fluorescence intensity, namely the gene expression intensity, between the RNAs derived from two types of cells (average difference) to calculate the fold changes thereof. Then, genes for which an increase or decrease more than 5-fold compared with at least one control specimen, or more than 3-fold compared with both of the two control specimens were selected (Table 2).

Table 2 Expression profile of the selected genes

chip set	selection method	3	fold →	5or13or18	← fold	12	13	fold→	14	description
AA004509	I5bI5c									C-MAMMA1002351
AA004509		−97	˜4.0	(18)213	˜8.5	−49				zh94f12.s1 Soares fetal liver spleen 1NFLS S1 Homo sapiens cDNA clone 428975 3 .
AA020825	I18bI13aI18 aI13cI18c									C-HEMBB1001294
AA020825		2	˜6.1	(13)76	˜4.1	17				ze64b02.s1 Soares retina N2b4HR Homo sapiens cDNA clone 363723 3 .
AA020825		2	˜9.7	(18)180	˜7.6	17				ze64b02.s1 Soares retina N2b4HR Homo sapiens cDNA clone 363723 3 .
AA027223	I13aI18aI5a									C-NT2RP2001327
AA027223		−37	˜8.4	(5)164						zk01a01.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 469224 3 .
AA027223		−37	˜11.2	(13)289						zk01a01.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 469224 3 .
AA027223		−37	˜6.4	(18)156						zk01a01.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 469224 3'.
AA113139	I5bI5aI5cI									C-NT2RM1000355
AA113139		22	˜7.7	(5)701	˜19.4	18				zm27a04.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone 526830 3 .
AA115259	I13bI13aI13 c									C-NT2RP2000459
AA115259		14	˜8.4	(5)201	6.6	30				zl08a10.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 491706 3 .
AA115562	I14									C-NT2RP2001420
AA115562							198	5.6	936	zl07c11.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 491636 3 .
AA126752	I5bI13aI5aI 5cI									C-Y79AA1000784 C-NT2RM4001382
AA126752		63	15	(5)874	6.4	110				zk95f03.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 490589 3 .
AA126752		63	˜11.8	(13)226						zk95f03.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 490589 3 .
AA127605	I5aI5c									C-NT2RM1000055
AA127605		34	5.3	(5)212	3.9	38				zn81f08.r1 Stratagene lung carcinoma 937218 Homo sapiens cDNA clone 564615 5 .
AA135406	I5bI5c									C-PLACE1008947
AA135406		−62	˜3.3	(5)190	˜6.2	−17				zo28e08.s1 Stratagene colon (#937204) Homo sapiens cDNA clone 588230 3 .
AA147884	I13bI18bI13									C-HEMBA1002716

	al18al13cl1 8c						
AA147884		−6	~5.9	(13)145	~11.2	11	zl50b04.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 505327 3 .
AA147884		−6	~5.0	(18)116	~7.0	11	zl50b04.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 505327 3 .
AA235118	l5bl5al5c						C-MAMMA1002461
AA235118		459	7.9	(5)2545	6	323	zs36f07.s1 Soares NhHMPu S1 Homo sapiens cDNA clone 687301 3 similar to contains element MSR1 repetitive element ;.
AA242823	l13bl13al13 c						C-NT2RP2002193
AA242823		−31 3	~14.1	(13)7	~8.8	−34	zr65e10.s1 Soares NhHMPu S1 Homo sapiens cDNA clone 668298 3 .
AA255525	l13bl13c						C-THYRO1000401
AA255525		66	3.9	(13)214	~7.9	−87	zr85a12.s1 Soares NhHMPu S1 Homo sapiens cDNA clone 682462 3 .
AA258267	l5c						C-NT2RP3004041
AA258267		10	~3.0	(5)66	~3.5	1	zr60h08.s1 Soares NhHMPu S1 Homo sapiens cDNA clone 667839 3 .
AA281528	l13bl13all13 c						C-OVARC1000781
AA281528		−91	~12.5	(13)225	~9.5	−18	zt08g09.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:712576 3 .
AA292158	l13al18al13 c						C-PLACE4000052
AA292158		2	~10.0	(13)319	3.3	97	zt46c03.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 725380 5 .
AA292158		2	~7.8	(18)112			zt46c03.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 725380 5'.
AA323430	l18b						C-NT2RP3002948
AA323430				(18)114	~6.2	−6	EST26202 Cerebellum II Homo sapiens cDNA 5' end similar to similar to ring canal protein.
AA378597	l13a						C-PLACE1001845,
AA378597		−24 6	~27.4	(13)559			EST91316 Synovial sarcoma Homo sapiens cDNA 5 end.
AA379742	l5al						C-NT2RM2001637
AA379742		−53	~8.0	(5)147			EST92623 Skin tumor I Homo sapiens cDNA 5 end.
AA398596	l13bl5bl13al 5al13cl5clS						C-PLACE1006469
AA398596		48	~13.3	(5)380	5.1	75	zt70a05.s1 Soares testis NHT Homo sapiens cDNA clone 727664 3 .
AA398596		48	~7.9	(13)153	~10.0	75	zt70a05.s1 Soares testis NHT Homo sapiens cDNA clone 727664 3 .
AA399226	l5bl						C-HEMBA1002417
AA399226				(5)170	~7.4	−1	zt50c01.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 725760 3 .
AA402715	l14l18a						C-PLACE1000786

AA402715		539	7.3	(18)3949			zu47c06.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 741130 3'.
AA402823	13b|18b						C-MAMMA1000416
AA402823				(13)146	~7.2	−125	zu55g07.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 741948 3 .
AA402823				(18)287	~8.7	−125	zu55g07.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 741948 3'.
AA410311	18a						C-PLACE1005409
AA410311		−13 8	~25.7	(18)615			zv23c07.s1 Soares NhHMPu S1 Homo sapiens cDNA clone 754476 3'.
AA410343	5a						C-HEMBB1002600
AA410343		−17 97	~29.7	(5)63			zv16e11.s1 Soares NhHMPu S1 Homo sapiens cDNA clone 753836 3 .
AA422049	18a						C-NT2RP3000605
AA422049		25	7.3	(18)200			zv28g05.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 755000 3' similar to gb:J02621 NONHISTONE CHROMOSOMAL PROTEIN HMG-14 (HUMAN);.
AA426218	13b|5b						C-NT2RM4002390
AA426218				(5)257	~8.5	12	zw17c11.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 769556 3 .
AA426218				(13)157	~5.4	12	zw17c11.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 769556 3 .
AA427861	13b|18b|13a|18a|13c|18c						C-HEMBA1004055
AA427861		68	10	(13)253	6.5	44	zw50b01.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone 773449 3 .
AA427861		68	5.2	(18)295	6.6	44	zw50b01.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone 773449 3 .
AA429917	13b						C-PLACE1005603
AA429917				(13)444	~21.5	−25	zw66f03.s1 Soares testis NHT Homo sapiens cDNA clone 781181 3 .
AA430355	18a|18c						C-HEMBA1002150
AA430355		151	7.6	(18)1227	3.4	366	zw20e04.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 769854 3 .
AA430674	13a|5a						C-Y79AA1000258
AA430674		−45	~19.5	(5)518			zw26d12.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 770423 3 .
AA430674		−45	~12.2	(13)297			zw26d12.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 770423 3 .
AA433899	13b						C-NT2RM1001105
AA433899				(13)141	~12.9	−47	zw52b06.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone 773651 3 .

AA445994	I5a									C-NT2RM4000027
AA445994		4	~6.5	(5)153						zw64e04.s1 Soares testis NHT Homo sapiens cDNA clone 780990 3 .
AA449773	I14I5a									C-MAMMA1002143
AA449773		86	9.2	(5)786						zx07h07.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone 785821 3 .
AA449773							77	13.1	978	zx07h07.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone 785821 3 .
AA453435	I18a									C-PLACE1006037
AA453435		94	6.4	(18)1292						zx32h03.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone 788213 3'.
AA453624	I5bI5c									C-NT2RP4000973
AA453624		89	3.4	(5)288	6.1	41				zx48c02.s1 Soares testis NHT Homo sapiens cDNA clone 795458 3 similar to gb:M11722 DNA NUCLEOTIDYLEXOTRANSFERASE (HUMAN);.
AA460708	I13bI13c									C-OVARC1001270
AA460708		84	3	(13)231	7	33				zx69e03.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone 796732 3 .
AA461093	I18bI13aI18aI13cI18c									C-HEMBB1001482
AA461093		-68	~5.6	(13)47	~3.6	-5				zx63f06.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone 796163 3 .
AA461093		-68	~8.6	(18)141	~6.5	-5				zx63f06.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone 796163 3 .
AA465367	I5a									C-NT2RP2005360
AA465367		-8	~6.4	(5)182						aa23d09.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:814097 3 .
AA478794	I13bI13c									C-MAMMA1000416
AA478794		-9	~4.5	(13)91	~7.2	1				zv20e01.s1 Soares NhHMPu S1 Homo sapiens cDNA clone 754200 3 .
AA489000	I13aI									C-PLACE1000133
										C-NT2RP2004013
AA489000		27	~5.3	(13)110						aa54d02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:824739 3 .
AA489080	I5aI5c									C-HEMBA1003615
AA489080		86	5.3	(5)455	4.5	100				aa54h08.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:824799 3 .
AA598982	I18b									C-PLACE3000242
AA598982				(18)246	~10.0	-50				ae34e01.s1 Gessler Wilms tumor Homo sapiens cDNA clone 897720 3' similar to contains element PTR5 repetitive element ;.
AA599674	I5bII5aII5cI									C-NT2RM2000522
										C-HEMBA1002475
										C-NT2RP2004242
AA599674		-18	~8.7	(5)750	12.7	59				ag10e11.s1 Gessler Wilms tumor Homo sapiens cDNA clone 1069964 3 .
AA620295	I5aII5c									C-NT2RM2001637

AA620295		16	˜10.2	(5)340	3.6	88	af04h10.s1 Soares testis NHT Homo sapiens cDNA clone 1030723 3 .
C02472	l5a						C-Y79AA1000784 C-NT2RM4001382
C02472		35	11	(5)454			HUMGS0012359, Human Gene Signature, 3 −directed cDNA sequence.
H49440	l13a						C-NT2RP3003290
H49440		56	˜11.3	(13)195			yo23d12.r1 Homo sapiens cDNA clone 178775 5 similar to contains Alu repetitive element;contains PTR7 repetitive element ;.
H61476	l5bl5c						C-HEMBA1004889
H61476		62	˜4.5	(5)439	˜11.1	55	yr17e08.s1 Homo sapiens cDNA clone 205574 3 .
N22273	l13al5a						C-HEMBA1006676 (HELIX) 2869 bp C-NT2RM2001696 (HELIX) 2661 bp
N22273		0	˜5.5	(5)238			
N22273		0	˜5.7	(13)184			
N30796	l5c						C-NT2RM4002593
N30796		66	˜4.6	(5)573	3.3	142	yw65d03.s1 Homo sapiens cDNA clone 257093 3 .
N31610	l5c						C-Y79AA1001781
N31610		−21 1	˜3.5	(5)73	˜3.4	−10	yy20g10.s1 Homo sapiens cDNA clone 271842 3 .
N39361	l5bl5c						C-HEMBA1003805
N39361		156	3.1	(5)109	˜5.9	−72	yx80d09.r1 Homo sapiens cDNA clone 268049 5 .
N40170	l5b						C-NT2RP2002606 C-NT2RP3003876
N40170				(5)130	˜5.2	−5	yy44b06.s1 Homo sapiens cDNA clone 276371 3 .
N73762	l13b						C-HEMBA1007085
N73762				(13)842	6	150	za61f08.s1 Homo sapiens cDNA clone 297063 3 .
N78718	l13a						C-NT2RP3000109
N78718		51	5.2	(13)280			zb02f10.s1 Homo sapiens cDNA clone 300907 3 .
R05274	l18b						C-PLACE1004506
R05274				(18)734	5.7	118	ye91b06.s1 Homo sapiens cDNA clone 125075 3'.
R06271	l18al18c						C-PLACE1005409
R06271		79	7.9	(18)881	4.1	180	yf08e02.s1 Homo sapiens cDNA clone 126266 3 .
R31785	l5bl5al5c						C-OVARC1001726
R31785		−91 3	˜15.1	(5)911	˜33.3	−555	yh68g11.s1 Homo sapiens cDNA clone 134948 3 .
R44761	l13a						C-NT2RP2003272
R44761		19	˜6.3	(13)471			yg30h03.s1 Homo sapiens cDNA clone 34148 3 similar to contains MER28 repetitive element ;.
R54743	l13bl5bl						C-HEMBA1005621
R54743				(5)492	13.5	36	yj75a07.r1 Homo sapiens cDNA clone 154548 5 .
R54743				(13)209	5.8	36	yj75a07.r1 Homo sapiens cDNA clone 154548 5 .
R56678	l13b						C-NT2RP3002399
R56678				(13)85	˜5.5	15	yi04d08.r1 Homo sapiens cDNA clone 138255 5 similar to contains Alu repetitive element;.
T10166	l13c						C-NT2RM2000101 C-NT2RP2002208
T10166		61	4.1	(13)249	4.2	94	seq879 Homo sapiens cDNA clone b4HB3MA-COT8-HAP-Ft1 66 3 .
T33018	l18al5al						C-NT2RM4000514

T33018		−26 3	~10.6	(5)407						EST56331 Homo sapiens cDNA 3 end similar to None.
T33018		−26 3	~6.1	(18)221						EST56331 Homo sapiens cDNA 3' end similar to None.
T47788	I5a									C-NT2RM1000039
T47788		−19 2	~6.6	(5)260						yb17a11.s1 Homo sapiens cDNA clone 71420 3 .
T64575	I5aII									C-MAMMA1001388
T64575		254	6.3	(5)1387						yc25a03.s1 Homo sapiens cDNA clone 81676 3 .
T71373	I5bII5aII5cI									C-MAMMA1001388
T71373		−54 5	~20.3	(5)251	~43.6	−775				yc61h07.s1 Homo sapiens cDNA clone 85213 3 .
T90699	I18bI18c									C-NT2RP3002273 C-MAMMA1000284
T90699		−93	~3.7	(18)234	~6.1	24				ye16d10.s1 Homo sapiens cDNA clone 117907 3 similar to contains PTR5 repetitive element ;.
T95057	I13bI5b									C-HEMBA1007085
T95057				(5)408	~5.4	25				ye39d04.s1 Homo sapiens cDNA clone 120103 3 .
T95057				(13)847	16.8	25				ye39d04.s1 Homo sapiens cDNA clone 120103 3 .
T97111	I5b									C-NT2RM2001345
T97111				(5)229	8.2	−38				ye41d04.r1 Homo sapiens cDNA clone 120295 5 .
T99474	I5c									C-NT2RP2000289
T99474		−9	~3.0	(5)223	3.2	70				ye64d12.s1 Homo sapiens cDNA clone 122519 3 .
W27237	I14									C-MAMMA1002143
W27237							31	12.1	444	24c11 Human retina cDNA randomly primed sublibrary Homo sapiens cDNA.
W68734	I5bI5aI5c									C-NT2RM4001155
W68734		−23 4	~6.6	(5)319	~11.3	−7				zd37f08.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone 342855 3 .
W72547	I13a									C-HEMBA1004669
W72547		36	6.2	(13)220						zd64g12.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone 345478 3 .
W86853	I5bI5c									C-NT2RP3002818
W86853		20	~3.8	(5)98	~5.6	−34				zh59d05.s1 Soares fetal liver spleen 1NFLS S1 Homo sapiens cDNA clone 416361 3 .
Z38501	I13bI18b									C-NT2RP3001730
Z38501				(13)144	~5.8	29				H. sapiens partial cDNA sequence; clone c-0de11.
Z38501				(18)141	~5.7	29				H. sapiens partial cDNA sequence; clone c-0de11.

A gene indicated "5" in the column of "selection method" in the table represents a gene identified by the expression analysis using the stomach cancer tissue (#5) transplanted to a SCID mouse, while a gene indicated "13"and "18" in the column of "selection method" represents a gene identified by the expression analysis using the clinical stomach cancer specimens (#13 and #18). Increases in the expression level of the gene in these three cancer tissues compared

with that in the two normal clinical specimens, #3 and #12 (#12 and #13 are the same specimen) are shown in this table. "a" indicates that the increase in the expression level (fold change) is more than 5-fold compared with the normal clinical specimen #3, while "b" shows that the increase in the expression level (fold change) is more than 5-fold compared with the normal clinical specimen #12. "C" indicates that the increase in the expression level (fold change) is more than 3-fold compared with both of the normal clinical specimens #3 and also #12. "14" represents a gene identified by the expression analysis using the lymph node metastasis of the stomach cancer clinical specimen #13, and which shows a "fold change" of more than 5-fold compared with that of #13. The expression amount (average difference; in the column of "5 or 13 or 18" of the table, the specimen numbers are shown in parenthesis) and fold change (in the table, two specimens compared to each other are shown with the sign "fold→" or "←fold") in each specimen are also shown in this table.

Additionally to these experiments, a similar experiment was performed also on the hepatoma. More specifically, expression levels of the genes were compared by the same differential analysis as above using a hepatic cancer tissue derived from a hepatitis B virus-infected patient (specimen #5) and a non-cancerous tissue (liver cirrhosis) derived from the same patient, which showed that the expression (average difference) of the above MAMMA1000416 was "55" and "569" in the non-cancerous tissue (liver cirrhosis) and the hepatic cancer tissue, respectively. That is, the ratio (fold change) of the expression level of this gene in the hepatic cancer tissue to that in the non-cancerous tissue (liver cirrhosis) is "~4.8", which indicates that the expression of MAMMA1000416 is also elevated in hepatoma.

2. Full length cDNA database

The NT-2 neuron progenitor cells (Stratagene), a teratocarcinoma cell line from human embryo testis, which can differentiate into neurons by the treatment with retinoic acid was used. The NT-2 cells were cultured according to the manufacturer's instructions as follows.

(1) NT-2 cells were cultured without induction by retinoic acid treatment (NT2RM1, NT2RM2, and NT2RM4).

(2) After cultured, NT-2 cells were induced by adding retinoic acid, and then were cultured for 2 weeks (NT2RP2, NT2RP3, and NT2RP4).

Also, the human retinoblastoma cell line Y79 (ATCC HTB-18) (Y79AA1) were cultured according to the culture conditions described in the ATCC catalogue (http://www.atcc.org/). The cells were harvested separately, and mRNA was extracted from cell by the method described in the literature (Molecular Cloning 2nd edition. (1989) Sambrook J., Fritsch, E.F., and Maniatis T., Cold Spring Harbor Laboratory Press). Furthermore, poly $(A)^+$RNA was purified from the mRNA using oligo-dT cellulose.

Similarly, human placenta (PLACE1, PLACE3, PLACE4), human ovary cancer tissue (OVARC1), tissues from human embryo at 10 weeks, which is enriched with head (HEMBA1), or body (HEMBB1), human mammary gland (MAMMA1) and human thyroid gland (THYRO1) were used to extract mRNA by the method described in the literature (Molecular Cloning 2nd edition. (1989) Sambrook J., Fritsch, E.F., and Maniatis T., Cold Spring Harbor Laboratory Press). Furthermore, poly $(A)^+$RNA was purified from the mRNA using oligo-dT cellulose.

Each poly $(A)^+$RNA was used to construct a cDNA library by the oligo-capping method (Maruyama M. and Sugano S. (1994) Gene, 138: 171-174). Using the Oligo-cap linker (agcaucgagu cggccuuguu ggccuacugg /SEQ ID NO: 150) and the Oligo-dT primer (gcggctgaag acggcctatg tggccttttt tttttttttt tt/ SEQ ID NO: 151), bacterial alkaline phosphatase (BAP) treatment, tobacco acid phosphatase (TAP) treatment, RNA ligation, the first strand cDNA synthesis, and removal of RNA were performed as described in the reference (Suzuki and Kanno, Protein Nucleic acid and Enzyme, 41: 197-201(1996); Suzuki Y. et al. Gene, 200: 149-156(1997)). Next, 5'- and 3'-PCR primers (agcatcgagt cggccttgtt g / SEQ ID NO: 152, and gcggctgaag acggcctatg t/ 153, respectively) were used for performing PCR to convert the cDNA into double stranded cDNA, which was then digested with SfiI. Then, the DraIII-cleaved pUC19FL3 vector (NT2RM1), or the DraIII- cleaved pME18SFL3 (GenBank AB009864, expression vector; for NT2RM2, NT2RM4, NT2RP2, NT2RP3, NT2RP4, Y79AA1, PLACE1, PLACE3, PLACE4, OVARC1, HEMBA1, HEMBB1, MAMMA1 and THYRO1) was used for cloning the cDNA in a unidirectional manner, and cDNA libraries were obtained. The

nucleotide sequence of the 5′- and 3′- ends of the cDNA clones was analyzed with a DNA sequencer (ABI PRISM 377, PE Biosystems) after sequencing reactions were performed with the DNA sequencing reagents (Dye Terminator Cycle Sequencing FS Ready Reaction Kit, dRhodamine Terminator Cycle Sequencing FS Ready Reaction Kit, or BigDye Terminator Cycle Sequencing FS Ready Reaction Kit, PE Biosystems), according to the instructions. The data were compiled into a database.

The full-length-enriched cDNA libraries except those for NT2RM1 was constructed using eukaryotic expression vector pME18SFL3. pME18SFL3 contains SRα promoter and SV40 small t intron in the upstream of the cloning site, and SV40 polyA added signal sequence site in the downstream. As the cloning site of pME18SFL3 has asymmetrical DraIII sites, and the ends of cDNA fragments contain SfiI sites complementary to the DraIII sites, the cloned cDNA fragments can be inserted into the downstream of the SRα promoter unidirectionally. Therefore, clones containing full-length cDNA can be expressed transiently by introducing the obtained plasmid directly into COS cell. Thus, the clones can be analyzed very easily in terms of the proteins that are the gene products of the clones, or in terms of the biological activities of the proteins.

Based on the determined 5′-end nucleotide sequence, the fullness of each clone was assessed using analytical results of the ATGpr and ESTiMateFL. The ATGpr is a program developed by A. A. Salamov, T. Nishikawa and M. B. Swindells, Helix Research Institute, to predict whether an ATG codon is a translation initiation codon from the sequential characteristics in the vicinity of the ATG codon. On the other hand, the ESTiMateFL is a method developed by T. Nishikawa and Ohta, Helix Research Institute, for selecting clones of cDNAs, which are predicted to be full-length cDNAs at a high possibility by comparing them with the 5′-end and 3′-end sequences of ESTs in the public databases.

Accordingly, clones of cDNAs that are predicted to be a full-length cDNA at a high possibility have been selected according to the fullness assessment. Furthermore, the 5′-end and 3′-end nucleotide sequences of these clones were searched in the public database to select clones that could be judged as novel.

Full-length sequence was determined for each selected cDNA clones. The nucleotide sequence determination was performed mainly by the dye-terminator method using custom synthesized DNA primers according to the primer walking procedure (custom synthesized DNA primers were used for sequencing; sequencing reaction was performed with DNA sequencing reagent supplied by PE Biosystems according to the supplier's manual; and the samples were analyzed in an automatic sequencer made by the same supplier). Sequence determination of some clones was carried out in the same manner but using a Licor DNA sequencer. Overlapping partial nucleotide sequences, which were obtained by the above-described method, were assembled together to determine a full-length nucleotide sequence. Amino acid sequences were then deduced from the determined full-length nucleotide sequences. The determined full-length nucleotide sequences and putative amino acid sequences were incorporated into a database to build a full-length cDNA database.

3. Collation with nucleotide sequences selected by the DD method

It was revealed that none of the sequences of the 76 clones selected in "1" is identical to known nucleotide sequences in the full-length cDNA database of "2" (i.e. they are novel), and further, that they comprise the same nucleotide sequences with the cDNA clones which had been judged to be full-length cDNA clones. SEQ ID NOs of the nucleotide sequences of the full-length cDNA clones, the nucleotide sequences of which were identical, and the corresponding amino acid sequences are shown in Table 1.

Finally, genes have been selected which are represented by sequences described in the column of the "selection method" in Table 2 as "13a" (five times or more in #13 compared with #3), "13b" (five times or more in #13 compared with #12), "13c" (three times or more in #13 compared with #3, and three times or more in #13 compared with #12), "18a" (five times or more in #18 compared with #3), "18b" (five times or more in #18 compared with #12), and "18c" (three times or more in #18 compared with #3, and three times or more in #18 compared with #12) as genes of which expression level increased in a cancerous tissue (#13 or #18) five times or more compared with those in the

normal gastric mucosa (#3 or #12), or three times or more compared with both of the normal gastric mucosa (#3 and #12). More specifically, such genes include the following: HEMBB1001294, NT2RP2001327, NT2RP2000459, Y79AA1000784, NT2RM4001382, HEMBA1002716, NT2RP2002193, THYRO1000401, OVARC1000781, PLACE4000052, NT2RP3002948, PLACE1001845, PLACE1006469, PLACE1000786, MAMMA1000416, PLACE1005409, NT2RP3000605, NT2RM4002390, HEMBA1004055, PLACE1005603, HEMBA1002150, Y79AA1000258, NT2RM1001105, PLACE1006037, OVARC1001270, HEMBB1001482, MAMMA1000416, PLACE1000133, NT2RP2004013, PLACE3000242, NT2RP3003290, HEMBA1006676, NT2RM2001696, HEMBA1007085, NT2RP3000109, PLACE1004506, PLACE1005409, NT2RP2003272, HEMBA1005621, NT2RP3002399, NT2RM2000101, NT2RP2002208, NT2RM4000514, NT2RP3002273, MAMMA1000284, HEMBA1007085, HEMBA1004669, and NT2RP3001730.

Furthermore, genes have been selected which are described as "14" in the column of "selection method" in Table 2 as genes of which expression level increased in the cancerous tissue of a lymph node metastatic focus (#14) five times or more compared with that in the stomach cancer tissue (#13). These genes include the following: NT2RP2001420, PLACE1000786, and MAMMA1002143.

Further, the following gene was selected as a gene of which expression level increased five times or more in the stomach cancer cell strain OCUM-2MD3, having a high peritoneal metastatic capability, compared with the stomach cancer cell strain OCUM-2M: MAMMA1001388.

Furthermore, genes have been selected which comprise sequences described in the column of selection method of Table 2 as "5a" (five times or more in #5 compared with #3), "5b" (five times or more in #5 compared with #12), or "5c" (three times or more in #5 compared with #3, and three times or more compared with #12) as genes of which expression level increased five times or more in the stomach cancer transplanted into nude mice (SCID) compared with the excised normal gastric mucosal cells (#3 or #12), or three times or more compared with both of the excised normal gastric mucosal cells (#3 and #12). These genes include the following: MAMMA1002351, NT2RP2001327, NT2RM1000355, Y79AA1000784, NT2RM4001382, NT2RM1000055,

PLACE1008947, MAMMA1002461, NT2RP3004041, NT2RM2001637, PLACE1006469, HEMBA1002417, HEMBB1002600, NT2RM4002390, Y79AA1000258, NT2RM4000027, MAMMA1002143, NT2RP4000973, NT2RP2005360, HEMBA1003615, NT2RM2000522, HEMBA1002475, NT2RP2004242, NT2RM2001637, Y79AA1000784, NT2RM4001382, HEMBA1004889, HEMBA1006676, NT2RM2001696, NT2RM4002593, Y79AA1001781, HEMBA1003805, NT2RP2002606, NT2RP3003876, OVARC1001726, HEMBA1005621, NT2RM4000514, NT2RM1000039, MAMMA1001388, MAMMA1001388, HEMBA1007085, NT2RM2001345, NT2RP2000289, NT2RM4001155, and NT2RP3002818.

4.Characteristics of the selected clones

The assessment results of the fullness for these clones using the ATGpr are shown in the following. The ATGpr, developed by Salamov A.A., Nishikawa T., and Swindells M.B. in the Helix Research Institute, is a program for prediction of the translation initiation codon based on the characteristics of the sequences in the vicinity of the ATG codon (A. A. Salamov, T. Nishikawa, M. B. Swindells, Bioinformatics, 14: 384-390 (1998); http://www.hri.co.jp/atgpr/). The results are shown with expectations (also described as ATGpr1 below) that an ATG is a true initiation codon (can be described as ATGpr1).

HEMBA1002150 0.31
HEMBA1002417 0.83
HEMBA1002475 0.88
HEMBA1002716 0.14
HEMBA1003615 0.94
HEMBA1003805 0.94
HEMBA1004055 0.74
HEMBA1004669 0.94
HEMBA1004889 0.94
HEMBA1005621 0.94
HEMBA1006676 0.17
HEMBA1007085 0.73
HEMBB1001294 0.86
HEMBB1001482 0.44
HEMBB1002600 0.91

MAMMA1000284	0.35	
MAMMA1000416	0.89	
MAMMA1001388	0.94	
MAMMA1002143	0.91	
MAMMA1002351	0.89	
MAMMA1002461	0.49	
NT2RM1000039	0.77	
NT2RM1000055	0.89	
NT2RM1000355	0.94	
NT2RM1001105	0.94	
NT2RM2000101	0.77	
NT2RM2000522	0.91	
NT2RM2001345	0.94	
NT2RM2001637	0.71	
NT2RM2001696	0.94	
NT2RM4000027	0.40	
NT2RM4000514	0.72	
NT2RM4001155	0.94	
NT2RM4001382	0.93	
NT2RM4002390	0.18	(maximum ATGpr2 was 0.24)
NT2RM4002593	0.91	
NT2RP2000289	0.06	(maximum ATGpr2 was 0.35)
NT2RP2000459	0.12	
NT2RP2001327	0.86	
NT2RP2001420	0.88	
NT2RP2002193	0.48	
NT2RP2002208	0.49	
NT2RP2002606	0.11	
NT2RP2003272	0.94	
NT2RP2004013	0.48	
NT2RP2004242	0.94	
NT2RP2005360	0.12	
NT2RP3000109	0.18	
NT2RP3000605	0.92	
NT2RP3001730	0.77	
NT2RP3002273	0.90	

NT2RP3002399	0.91
NT2RP3002818	0.91
NT2RP3002948	0.60
NT2RP3003290	0.62
NT2RP3003876	0.42
NT2RP3004041	0.52
NT2RP4000973	0.36
OVARC1000781	0.80
OVARC1001270	0.48
OVARC1001726	0.18
PLACE1000133	0.53
PLACE1000786	0.88
PLACE1001845	0.08
PLACE1004506	
PLACE1005409	0.09
PLACE1005603	0.92
PLACE1006037	0.65
PLACE1006469	0.85
PLACE1008947	0.05
PLACE3000242	0.94
PLACE4000052	0.80
THYRO1000401	0.73
Y79AA1000258	0.36
Y79AA1000784	0.93
Y79AA1001781	0.74

Next, it was predicted whether the amino acid sequences deduced from these clones comprise a signal sequence in the amino-terminus and transmembrane regions. Furthermore, protein functional domain (motif) analysis was performed. The examinations for a signal sequence in the amino-terminus, for a transmembrane region and for a functional domain were performed by using PSORT [K. Nakai & M. Kanehisa, Genomics, 14:897-911 (1992)], SOSUI [T. Hirokawa et al., Bioinformatics, 14:378-379 (1998)] (Mitsui Knowledge Industry Co., Ltd.) and Pfam (http://www.sanger.ac.uk/Software/Pfam/index.shtml), respectively. When the presence of a signal sequence or a transmembrane region in

the amino-terminus was predicted in the amino acid sequence by PSORT or SOSUI, the protein was predicted to be a secretory protein or a membrane protein. When the amino acid sequence matched a functional domain in the Pfam search for a functional domain, the function of the protein is predictable based on the matching data, for example, by referring to the functional categories in PROSITE (http://www.expasy.ch/cgi-bin/prosite-list.pl). The functional domain search can be also performed by using PROSITE.

As a result, a signal sequence was detected in the putative amino acid sequence of Y79AA1000258 by PROST. Furthermore, transmembrane regions were detected using SOSUI in the putative amino acid sequences of HEMBA1002150, HEMBA1004889, HEMBB1002600, MAMMA1000416, MAMMA1001388, MAMMA1002461, NT2RM1000355, NT2RP2000289, NT2RP2000459, NT2RP4000973, PLACE4000052, HEMBA1004055, and Y79AA1000258.

In the following, the results of homology search against known gene databases in the field based on the full-length nucleotide sequences and putative amino acid sequences of respective clones are shown. Each data has been described in the order of the sequence nomenclature, definition of a hit data of the highest analogy, P value, length of compared sequence, homology, and accession number of the hit data divided by the sign "//". Herein, P value is a score of the similarity among sequences taking the statistical probability of its occurrence into consideration; in general, the smaller the P value, the higher the analogy will be (Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & amp; Lipman, D. J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215: 403-410; Gish, W. & States, D. J. (1993) "Identification of protein coding regions by database similarity search." Nature Genet. 3: 266-272).

HEMBA1002417//"Homo sapiens chromosome 19, cosmid R28784, complete sequence."//1.4E-299//294bp//100%//AC005954
HEMBA1002417//TIGHT JUNCTION PROTEIN ZO-1 (TIGHT JUNCTION PROTEIN 1).//1.00E-121//489aa//52%//P39447
HEMBA1002475//SKIN SECRETORY PROTEIN XP2 PRECURSOR (APEG PROTEIN).//1.10E-12//285aa//31%//P17437

HEMBA1003615//Homo sapiens ART-4 mRNA, complete cds.//0//1713bp//99%//AB026125
HEMBA1003805//Mus musculus KH domain RNA binding protein QKI-5A mRNA, complete cds.//0//988bp//95%//AF090402
HEMBA1004669//SON PROTEIN (SON3).//7.30E-17//288aa//36%//P18583
HEMBA1004889//Human C3f mRNA, complete cds.//6.70E-24//341aabp//26%//U72515
HEMBA1005621//"Homo sapiens Mad2B protein (MAD2B) mRNA, complete cds."//2.9E-224//1031bp//99%//AF139365
HEMBA1005621//Homo sapiens Mad2-like protein mRNA, complete cds.//8.00E-211//962bp//99%//AF072933
HEMBB1001294//GTP-BINDING PROTEIN TC10.//1.20E-79//196aa//80%//P17081
HEMBB1001482//ZINC FINGER PROTEIN 91 (ZINC FINGER PROTEIN HTF10) (HPF7).//2.10E-57//941aa//27%//Q05481
HEMBB1002600//Homo sapiens tetraspan NET-5 mRNA, complete cds.//0//1417bp//99%//AF089749
MAMMA1000284//P.walti mRNA for rnp associated protein 55.//2.20E-109//864bp//76%//X99836
MAMMA1000416//HYPOTHETICAL 32.0 KD PROTEIN C09F5.2 IN CHROMOSOME III.//2.00E-30//119aa//53%//Q09232
MAMMA1001388//LEUCINE-RICH ALPHA-2-GLYCOPROTEIN (LRG).//1.40E-165//312aa//99%//P02750
MAMMA1002143//Homo sapiens Cdc42 effector protein 4 mRNA, complete cds.//1.70E-252//1170bp//99%//AF099664
MAMMA1002351//FERRIPYOCHELIN BINDING PROTEIN.//0.000078//127aa//26%//P40882
MAMMA1002351//Mus musculus dynactin subunit p25 (p25) mRNA, complete cds.//4.30E-119//773bp//86%//AF190795
NT2RM1000039//HYPOTHETICAL 41.4 KD PROTEIN IN SRLQ-HYPF INTERGENIC REGION (EC 1.18.1.-) (ORF4) (ORF2).//2.90E-14//299aa//25%//P37596
NT2RM1000055//"Homo sapiens mRNA for KIAA0829 protein, partial cds."//0//3111bp//99%//AB020636
NT2RM1000055//Rattus norvegicus mRNA for TIP120, complete cds.//0//3106bp//89%//D87671
NT2RM1000355//Homo sapiens transmembrane protein BRI (BRI) mRNA,

complete cds.//0//1599bp//99%//AF152462
NT2RM2000522//SKIN SECRETORY PROTEIN XP2 PRECURSOR (APEG PROTEIN).//1.30E-12//282aa//32%//P17437
NT2RM2001345//VEGETATIBLE INCOMPATIBILITY PROTEIN HET-E-1.//2.90E-08//334aa//22%//Q00808
NT2RM4001155//ADRENAL MEDULLA 50 KD PROTEIN.//4.10E-197//445aa//78%//Q27969
NT2RM4001382//Homo sapiens RanBP7/importin 7 mRNA, complete cds.//2.20E-237//1079bp//99%//AF098799
NT2RP2001327//TUMOR NECROSIS FACTOR, ALPHA-INDUCED PROTEIN 1, ENDOTHELIAL (B12 PROTEIN).//5.50E-116//311aa//71%//Q13829
NT2RP2001420//Mus musculus nuclear protein NIP45 mRNA, complete cds.//9.00E-112//742bp//82%//U76759
NT2RP2002193//Homo sapiens PIAS3 mRNA for protein inhibitor of activated STAT3, complete cds.//0//2809bp//99%//AB021868
NT2RP2002606//Rattus norvegicus Rabin3 mRNA, complete cds.//9.20E-147//874bp//87%//U19181
NT2RP2003272//Homo sapiens ubiquilin mRNA, complete cds.//0//1789bp//99%//AF176069
NT2RP2004013//TRANSCRIPTION FACTOR BTF3 (RNA POLYMERASE B TRANSCRIPTION FACTOR 3).//2.30E-53//141aa//78%//P20290
NT2RP2004242//NEUROFILAMENT TRIPLET H PROTEIN (200 KD NEUROFILAMENT PROTEIN) (NF-H).//9.90E-12//427aa//26%//P19246
NT2RP2005360//Homo sapiens sentrin/SUMO-specific protease (SENP1) mRNA, complete cds.//1.30E-52//753bp//67%//AF149770
NT2RP3000109//P54 PROTEIN PRECURSOR.//0.0000065//358aa//22%//P13692
NT2RP3000605//Mus musculus mRNA for wizL, complete cds.//0//2232bp//82%//AB012265
NT2RP3001730//SEPTIN 2 HOMOLOG (FRAGMENT).//7.10E-132//294aa//84%//Q14141
NT2RP3002273//SCD6 PROTEIN.//1.30E-09//295aa//28%//P45978
NT2RP3002399//DNA REPLICATION LICENSING FACTOR MCM4 (CDC21 HOMOLOG) (P1-CDC21).//8.60E-79//416aa//34%//P33991
NT2RP3002818//INSERTION ELEMENT IS2A HYPOTHETICAL 48.2 KD PROTEIN.//5.70E-226//303aa//97%//P51026

NT2RP3002948//RING CANAL PROTEIN (KELCH PROTEIN).//2.00E-111//551aa//42%//Q04652
NT2RP3003290//Mus musculus mRNA for Ndr1 related protein Ndr3, complete cds.//1.5e-310//1468bp//82%//AB033922
NT2RP3003876//Rattus norvegicus Rabin3 mRNA, complete cds.//4.50E-147//874bp//87%//U19181
NT2RP4000973//PROBABLE PROTEIN DISULFIDE ISOMERASE P5 PRECURSOR (EC 5.3.4.1).//1.40E-26//90aa//42%//P38660
OVARC1001726//APICAL-LIKE PROTEIN (APXL PROTEIN).//4.30E-16//116aa//43%//Q13796
PLACE1000133//TRANSCRIPTION FACTOR BTF3 (RNA POLYMERASE B TRANSCRIPTION FACTOR 3).//1.80E-62//158aa//81%//P20290
PLACE1000786//PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF) (FACIOGENITAL DYSPLASIA PROTEIN HOMOLOG).//7.10E-09//59aa//47%//P52734
PLACE1001845//Mus musculus cyclin ania-6a mRNA, complete cds.//3.30E-31//925bp//62%//AF159159
PLACE1004506//Homo sapiens carboxyl terminal LIM domain protein (CLIM1) mRNA, complete cds.//2.10E-16//402bp//62%//U90878
PLACE1006469//ACETYL-COENZYME A SYNTHETASE (EC 6.2.1.1) (ACETATE--COA LIGASE) (ACYL- ACTIVATING ENZYME).//1.20E-83//313aa//49%//P27550
PLACE3000242//"Homo sapiens mRNA for KIAA1114 protein, complete cds."//0//2786bp//96%//AB029037
PLACE3000242//Human trophinin mRNA, complete cds.//0//2290bp//99%//U04811
PLACE4000052//Homo sapiens ATP cassette binding transporter 1 (ABC1) mRNA, complete cds.//0//4661bp//99%//AF165281
THYRO1000401//Human TcD37 homolog (HTcD37) mRNA, partial cds.//1.10E-90//430bp//99%//U67085
Y79AA1000784//"Homo sapiens RanBP7/importin 7 mRNA, complete cds."//1.10E-236//1076bp//99%//AF098799

5. Gene expression analysis with hybridization using high density DNA filter

Nylon membrane for DNA spotting was prepared according to the

following procedure. *E. coli* was cultured in each well of a 96-well plate (in a LB medium at 37°C for 16 hours). A small aliquot of each culture was suspended in 10 μl of sterile water in a well of a 96-well plate. The plate was heated at 100°C for 10 minutes. Then the boiled samples were analyzed by PCR reaction. PCR was performed in a 20 μl solution by using TaKaRa PCR Amplification Kit (Takara) according to the supplier's protocol. Primers used for the amplification of an insert cDNA in a plasmid were a pair of sequencing primers, ME761FW (5' tacggaagtgttacttctgc 3'/SEQ ID NO: 154) and ME1250RV (5' tgtgggaggtttttctcta 3'/SEQ ID NO: 155), or a pair of primers, M13M4 (5' gttttcccagtcacgac 3'/SEQ ID NO: 156) and M13RV (5' caggaaacagctatgac 3'/ SEQ ID NO: 157). PCR reaction was performed in a thermal cycler, GeneAmp System 9600 (PE Biosystems). The cycling profile consisted of pre-heat at 95°C for 5 minutes; 10 cycles of denaturation at 95°C for 10 seconds, and annealing/extension at 68°C for 1 minute; 20 cycles of denaturation at 98°C for 20 seconds and annealing/extension at 60°C for 3 minutes; and final extension at 72°C for 10 minutes. After the PCR reaction, the 20 μl reaction solution was loaded onto a 1% agarose gel and fractionated by electrophoresis. DNA on the gel was stained with ethidium bromide to confirm the amplification of cDNA. When cDNAs were barely amplified by PCR, plasmids containing the corresponding insert cDNAs were prepared by the alkali-extraction method (J. Sambrook, E.F., Fritsh, & T. Maniatis, "Molecular Cloning, A laboratory manual/ 2nd edition, Cold Spring Harbor Laboratory Press, 1989).

Preparation of DNA array was carried out by the following procedure. An Aliquot of a DNA solution was added in each well of a 384-well plate. DNA was spotted onto a nylon membrane (Boehringer) by using a 384-pin tool of Biomek 2000 Laboratory Automation System (Beckman-Coulter). Specifically, the 384-well plate containing the DNA was placed under the 384-pin tool. The independent 384 needles were simultaneously dipped into the DNA solution for DNA deposition. The needles were gently pressed onto a nylon membrane and the DNA deposited at the tips of needles was spotted onto the membrane. Denaturation of the spotted DNA and immobilization of the DNA on the nylon membrane were carried out according to usual manners (J. Sambrook,

E.F., Fritsh, & T. Maniatis, "Molecular Cloning, A laboratory manual/ 2nd edition, Cold Spring Harbor Laboratory Press, 1989).

Hybridization probe used was radioisotope-labeled 1st strand cDNA. The 1st strand cDNA synthesis was performed by using Thermoscript(TM) RT-PCR System (GIBCO). Specifically, the 1st strand cDNA was synthesized by using 1.5 μg mRNAs from various human tissues (Clontech), 1 μl aliquots of 50 μM Oligo (dT) 20 and 50 μCi [α^{33}P]dATP according to an attached protocol. Probe purification was carried out by using ProbeQuant(TM) G-50 micro column (Amersham-Pharmacia Biotech) according to an attached protocol. In the next step, 2 units of *E. coli* RNase H were added to the reaction mixture. The mixture was incubated at room temperature for 10 minutes and then 100 μg of human COT-1 DNA (GIBCO) was added thereto. The mixture was incubated at 97°C for 10 minutes and then was allowed to stand on ice to give hybridization probe.

Hybridization of the radioisotope-labeled probe to the DNA array was performed in a usual manner (J. Sambrook, E.F., Fritsh, & T. Maniatis, Molecular Cloning, A laboratory manual/ 2nd edition, Cold Spring Harbor Laboratory Press, 1989). The membrane was washed as follows: the nylon membrane was incubated in Washing solution 1 (2x SSC, 1% SDS) at room temperature (about 26°C) for 20 minutes and this washing was repeated 3 times; then the membrane was washed 3 times by incubating it in Washing solution 2 (0.1x SSC, 1% SDS) at 65°C for 20 minutes. Autoradiography was performed by using an image plate for BAS2000 (Fuji Photo Film). Specifically, the nylon membrane with probe hybridized thereon was wrapped with a piece of Saran Wrap and brought into tight contact with the image plate on the light-sensitive surface. The membrane with the image plate was placed in an imaging cassette for radioisotope and allowed to stand in dark place for 4 hours. The radioactivity recorded on the image plate was analyzed by using BAS2000. The activity was subjected to electronic conversion and recorded as an image file of autoradiogram. The signal intensity of each DNA spot was analyzed by using Visage High Density Grid Analysis Systems (Genomic Solutions). The signal intensity was converted into numerical data. The data were taken by duplicated measurements. The reproducibility was assessed by comparing the signal intensities of

the corresponding spots on the duplicated DNA filters that were hybridized to a single DNA probe. The ratio between the corresponding spots falls within a range of 2 or less in 95% of entire spots and the correlation coefficient is r=0.97. Thus the reproducibility is assumed to be satisfactory.

The detection sensitivity in gene expression analysis was estimated by examining increases in the signal intensity of probe concentration-dependent spot in hybridization using a probe complementary to the DNA spotted on the nylon membrane. DNA used was PLACE1008092 (same as DNA deposited in GenBank under an Accession No. AF107253). The DNA array with DNA of PLACE1008092 was prepared according to the above-mentioned method. The probe used was prepared as follows: mRNA was synthesized *in vitro* from the clone, PLACE1008092. By using this mRNA as a template, radioisotope-labeled 1st strand cDNA was synthesized in the same manner as described above, and the cDNA was used as the probe. The cDNA PLACE1008092 was inserted into pBluescript SK (-), of which T7 promoter was ligated to the 5'- end of the cDNA, to give a plasmid for *in vitro* synthesis of the mRNA from PLACE1008092. Specifically, the PLACE1008092 insert was cut out from pME18SFL3 carrying the cDNA at a DraIII site thereof by XhoI digestion. The resulting PLACE1008092 fragment was ligated to XhoI-predigested pBluescript SK (-) by using DNA ligation kit ver.2 (Takara). The *in vitro* mRNA synthesis from PLACE1008092 inserted in pBluescript SK (-) was carried out by using Ampliscribe(TM) T7 high yield transcription kit (Epicentre technologies). Hybridization and the analysis of signal intensity of each DNA spot were conducted by the same methods as described above. When the probe concentration is 1 x 10^7 μg/ml or less, there was no increase of signal intensity proportional to the probe concentration. Therefore it was assumed to be difficult to compare the signals with one another in the concentration range. Thus the spots with the intensity of 40 or less were indiscriminately taken as low-level signals. Within a concentration of the probe ranging from 1 x 10^7 μg/ml to 0.1 μg/ml, the signal was found to increase in a probe concentration-dependent manner. The detection limit was then assumed to be 1:100,000 in a ratio of mRNA expression level in a sample.

The expression levels of each cDNA in human normal tissues (heart, lung, pituitary gland, thymus, brain, kidney, liver and spleen) are indicated by numerical values of 0-10,000. As a result, genes that were expressed in at least a single tissue are indicated below by the corresponding clone names: HEMBA1002150, HEMBA1002417, HEMBA1003615, HEMBA1003805, HEMBA1004669, HEMBA1006676, HEMBA1007085, HEMBB1001294, MAMMA1000284, MAMMA1000416, MAMMA1001388, MAMMA1002143, MAMMA1002351, MAMMA1002461, NT2RM1000039, NT2RM1000355, NT2RM2000101, NT2RM2001345, NT2RM2001696, NT2RM4001155, NT2RM4001382, NT2RM4002593, NT2RP2000289, NT2RP2000459, NT2RP2001327, NT2RP2001420, NT2RP2002193, NT2RP2002208, NT2RP2003272, NT2RP2004013, NT2RP2005360, NT2RP3001730, NT2RP3002273, NT2RP3002399, NT2RP3003290, NT2RP3003876, OVARC1001726, PLACE1000786, PLACE1004506, PLACE1005409, PLACE1006469, PLACE1008947, PLACE3000242, PLACE4000052, THYRO1000401, and Y79AA1000258.

Genes that were expressed in all the tissues tested are indicated below by the corresponding clone names: HEMBA1002150, HEMBA1007085, MAMMA1000416, MAMMA1001388, and NT2RM1000039.

Genes that were expressed at low levels in any of the tissues tested are indicated below by the corresponding clone names: HEMBA1002475, HEMBA1002716, HEMBA1004055, HEMBA1004889, HEMBA1005621, HEMBB1001482, HEMBB1002600, NT2RM1000055, NT2RM1001105, NT2RM2000522, NT2RM2001637, NT2RM4000027, NT2RM4000514, NT2RM4002390, NT2RP2002606, NT2RP2004242, NT2RP3000109, NT2RP3000605, NT2RP3002818, NT2RP3002948, NT2RP3004041, NT2RP4000973, OVARC1000781, OVARC1001270, PLACE1000133, PLACE1001845, PLACE1005603, PLACE1006037, Y79AA1000784, and Y79AA1001781.

Genes exhibiting characteristic features in the expression thereof were selected by statistical analysis of these data. Examples are shown below to describe the selection of genes of which expression is varied greatly among tissues.

Gene of OVARC1000037 (heterogeneous nuclear ribonucleoprotein

(hnRNP)} which expression is varied little. Genes of which expression is varied greatly among tissues as compared that of the OVARC1000037 gene were determined as follows. Specifically, sum of squared deviation was calculated in the signal intensity observed in each tissue, which was divided by 7 degrees of freedom to determine a variance S_a^2. Next, sum of squared deviation was calculated in the signal intensity of a gene to be compared observed in each tissue, which was divided by 7 degrees of freedom to determine a variance S_b^2. By taking variance ratio F as $F=S_b^2/S_a^2$, genes with a significance level of 5% or more were extracted in the F distribution. As a result, HEMBA1002150, MAMMA1000416, NT2RM1000039 and NT2RM1000355 were extracted. Genes exhibiting characteristic features in the expression thereof were determined by comparing and statistical analysis of expression of numerous genes described above.

6. Analysis of disease-associated genes

Non-enzymic protein glycation reaction is believed to be a cause of a variety of chronic diabetic complications. Accordingly, genes of which expression is elevated or decreased in a glycated protein-specific manner in the endothelial cells are associated with diabetic complications caused by glycated proteins. Vascular endothelial cells are affected with glycated proteins present in blood. Reaction products of non-enzymic protein glycation include amadori compound (glycated protein) as a mildly glycated protein and advanced glycation endproduct as a heavily glycated protein. Hence, a survey was carried out for genes of which expression levels are varied depending on the presence of these glycated proteins in endothelial cells. The mRNAs were extracted from endothelial cells that were cultured in the presence or absence of glycated protein. The mRNAs were converted into radiolabeled first strand cDNAs for preparing probes. The probes were hybridized to the above-mentioned DNA array. Signal of each DNA spot was detected by BAS2000 and analyzed by ArrayGauge (Fuji Photo Film).

Advanced glycation endproduct of bovine serum albumin was prepared as follows: bovine serum albumin (BSA; Sigma) was incubated in a phosphate buffer solution containing 50 mM glucose at 37°C for

8 weeks; and the resulting brownish BSA was dialyzed against a phosphate buffer solution.

Human normal pulmonary arterial endothelial cells (Cell Applications) were cultured in an Endothelial Cell Growth Medium (Cell Applications). The culture dish (Falcon) with the cells was incubated in a CO_2 incubator (37°C, 5% CO_2, in a humid atmosphere). When the cells were grown to be confluent in the dish, 250 μg/ml of bovine serum albumin (Sigma), glycated bovine serum albumin (Sigma) or advanced glycation endproduct of bovine serum albumin was added thereto and the cells were incubated for 33 hours. The mRNA was extracted from the cells by using a FastTrack(TM) 2.0 kit (Invitrogen). The labeling of hybridization probe was carried out by using the mRNA according to the same procedure as described above.

As a result of measurement of the expression level of each cDNA in human pulmonary arterial endothelial cells cultured in a medium containing bovine serum albumin (sigma), glycated bovine serum albumin (Sigma) or advanced glycation endproduct of bovine serum albumin, genes of which expression was detected in the endothelial cell are as follows:
HEMBA1003615, HEMBA1003805, HEMBA1004669, HEMBA1007085, HEMBB1001294, HEMBB1002600, MAMMA1000284, MAMMA1000416, MAMMA1001388, MAMMA1002461, NT2RM1000039, NT2RM1000355, NT2RM2000101, NT2RM2001345, NT2RM2001696, NT2RM4000514, NT2RM4001382, NT2RP2001327, NT2RP2001420, NT2RP2002208, NT2RP2002606, NT2RP2003272, NT2RP2004013, NT2RP2004242, NT2RP2005360, NT2RP3001730, NT2RP3002273, NT2RP3002399, NT2RP3003290, NT2RP3003876, NT2RP3004041, NT2RP4000973, PLACE1000133, PLACE1001845, PLACE1004506, PLACE3000242, and Y79AA1000784.

7. Analysis of genes associated with neural cell differentiation

Genes involved in neural cell differentiation are useful for treating neurological diseases. It is possible that genes with varying expression levels in response to induction of cellular differentiation in neural cells are associated with neurological diseases.

A survey was performed for genes of which expression levels are

varied in response to induction of differentiation (stimulation by retinoic acid (RA)) in cultured cells of a neural strain, NT2.

The NT2 cells were treated basically according to supplier's instruction manual. "Undifferentiated NT2 cells" means NT2 cells successively cultured in an OPTI-MEM I (GIBCO-BRL; catalog No. 31985) containing 10%(v/v) fetal bovine serum and 1%(v/v) penicillin-streptomycin (GIBCO BRL). "NT2 cells cultured in the presence of retinoic acid," means the cells resulted from transferring undifferentiated NT2 cells into a retinoic acid-containing medium, which consists of D-MEM (GIBCO BRL; catalog No. 11965), 10%(v/v) fetal bovine serum, 1%(v/v) penicillin-streptomycin and 10 μM retinoic acid (GIBCO-BRL), and the subsequent successive culture therein for 5 weeks. "NT2 cells that were cultured in the presence of retinoic acid and then further cultured in the presence of cell-division inhibitor added" means NT2 cells resulted from transferring NT2 cells cultured in the presence of retinoic acid for 5 weeks into a cell-division inhibitor-containing medium, which consisted of D-MEM (GIBCO BRL; catalog No.11965), 10%(v/v) fetal bovine serum, 1%(v/v) penicillin-streptomycin, 10 μM retinoic acid, 10 μM FudR (5-fluoro-2'-deoxyuridine: GIBCO BRL), 10 μM Urd (Uridine: GIBCO BRL) and 1 μM araC (Cytosine β-D-Arabinofuranoside: GIBCO BRL), and the subsequence successive culture for 2 weeks. Each of the cells were treated with trypsin and then harvested. Total RNAs were extracted from the cells by using S.N.A.P.(TM) total RNA Isolation kit (Invitrogen). The labeling of probe used for hybridization was carried out by using 10 μg of the total RNA according to the same methods as described above.

The data were obtained in triplicate (n=3). The data of signal value representing gene expression level in the cells in the presence of stimulation for inducing differentiation were compared with those in the absence of the stimulation. The comparison was performed by statistical treatment of two-sample t-test. Clones with significant difference in the signal distribution were selected under the condition of $p < 0.05$. In this analysis, clones with the difference can be statistically detected even when the signals were low. Accordingly, clones with signal value of 40 or less were also assessed for the

selection.

Averaged signal values (M_1, M_2) and sample variances (s_1^2, s_2^2) were calculated for each gene in each of the cells, and then, the pooled sample variances s^2 were obtained from the sample variances of the two types of cells to be compared. The t values were determined according to the following formula: $t=(M1-M2)/s/(1/3+1/3)^{1/2}$. When the determined t-value was greater than a t-value at P, which means the probability of significance level, of 0.05 or 0.01 in the t-distribution table with 4 degrees of freedom, the difference was judged to be found in the expression level of the gene between the two types of cells at $p<0.05$ or $p<0.01$, respectively.

Expressions of HEMBA1003805, HEMBA1004669, HEMBA1007085, NT2RM1000039, NT2RM1001105, NT2RM2001637, NT2RP2001420, NT2RP2002193, NT2RP2002208, NT2RP2003272, NT2RP3000109, NT2RP3000605, NT2RP3003290, NT2RP3004041, PLACE1001845, PLACE1005409, and PLACE3000242 increased by RA. Expressions of NT2RM1000355, NT2RP2002193, NT2RP2003272, NT2RP3004041, PLACE1004506, PLACE1005603, and PLACE3000242 were elevated by RA/inhibitor. In addition, the expression of NT2RM4002593 decreased by RA/inhibitor. Furthermore, expressions of NT2RP2002193, NT2RP2003272, NT2RP3004041, and PLACE3000242 increased by both RA and RA/inhibitor. These clones are clones associated with neurodisorders.

8. Analysis of rheumatoid arthritis-associated genes

The onset of rheumatoid arthritis is thought to be involved in the proliferation of synovial cells covering inner surfaces of joint cavity and in inflammatory reaction resulted from the action of cytokines produced by leukocytes infiltrating into the joint synovial tissues (Rheumatism Information Center, http://www.rheuma-net.or.jp/). Recent studies have also revealed that tissue necrosis factor (TNF)-alpha participates in the onset (Current opinion in immunology 1999, 11, 657-662). When the expression of a gene exhibits responsiveness to the action of TNF on synovial cells, the gene is considered to be involved in rheumatoid arthritis.

A survey was performed for genes of which expression levels are

varied in response to TNF-alpha in the primary cell culture of synovial tissue. The primary cultured cells of the smooth muscle (Cell Applications) were grown to be confluent in a culture dish, and then, human TNF-alpha (Boehringer-Mannheim) was added at a final concentration of 10 ng/ml thereto. The culture was further continued for 24 hours.

Total RNA was extracted from the cells by using S.N.A.P.(TM) total RNA Isolation kit (Invitrogen). The labeling of probe used for hybridization was carried out by using 10 μg of the total RNA according to the same methods as described above. The data were obtained in triplicate (n=3). The data of signal value representing gene expression level in the cells in the presence of TNF stimulation were compared with those in the absence of the stimulation. The comparison was performed by statistical treatment of two-sample t-test. Clones with significant difference in the signal distribution were selected under the condition of $p<0.05$. In this analysis, clones with the difference can be statistically detected even when the signals were low. Therefore, clones with signal value of 40 or less were also assessed for the selection.

Averaged signal values (M_1, M_2) and sample variances (s_1^2, s_2^2) for each gene were calculated in each of the cells, and then, the pooled sample variances s^2 were obtained from the sample variances of the two types of cells to be compared. The t-values were determined according to the following formula: $t=(M_1-M_2)/s/(1/3+1/3)^{1/2}$. When the determined t-value was greater than a t-value at P, which means the probability of significance level, of 0.05 or 0.01 in the t-distribution table with 4 degrees of freedom, the difference was judged to be found in the expression level of the gene between the two types of cells at $p<0.05$ or $p<0.01$, respectively.

As a result, expression levels of HEMBA1004889, MAMMA1000416, NT2RM1000039, NT2RM2000101, NT2RM4000514, NT2RP2003272, NT2RP3002399, and Y79AA1000784 were elevated by TNF-alpha, while expression levels of HEMBA1002150, NT2RP3003290, and OVARC1001270 were reduced by TNF-alpha. These clones are associated with rheumatoid arthritis (rheumatism).

9. Analysis of ultraviolet radiation damage-associated genes

It is known that ultraviolet rays give considerably adverse influence on the health. In recent years, there have been significant risks of tissue damage by ultraviolet rays because of destruction of the ozone layer. Thus, ultraviolet radiation has been recognized as a risk factor for skin diseases such as skin cancers (United States Environmental Protection Agency: Ozone Depletion Home Page, http://www.epa.gov/ozone/). Genes of which expression levels are varied in skin epidermal cells exposed to ultraviolet rays are considered to be associated with skin damage caused by ultraviolet radiation.

After primary cultured skin fibroblast cells were irradiated with ultraviolet ray and were cultured, a survey was performed for genes of which expression levels were varied depending on the irradiation of ultraviolet ray. First, after cultured to be confluent, the primary cultured skin fibroblast cells (Cell Applications) were exposed to 10,000 $\mu J/cm^2$ of 254-nm ultraviolet light. Messenger RNAs were, then, extracted by using a FastTrack™ 2.0 mRNA Isolation kit (Invitrogen) from the unexposed cells and from the cells that were exposed to the ultraviolet light and then cultured for 4 or 24 hours. The labeling of the hybridization probe was carried out by using a 1.5 μg of each mRNA in the same manner as described above. The data were obtained in triplicate (n=3). The hybridization signals were compared between the cells exposed to the ultraviolet light and the unexposed cells. The comparison was preformed by statistical treatment with two-sample t-test. Clones with significant differences in the signal distribution were selected under the condition of $p<0.05$. In this analysis, even when the signal is lower than others, the difference in the signal values can be detected statistically. Accordingly, clones with signal value of 40 or lower were also assessed for selection.

Averaged signal values (M_1, M_2) and sample variances (s_1^2, s_2^2) were calculated for each gene in each of the cells, and then, the pooled sample variances s^2 were obtained from the sample variances of the two types of cells to be compared. The t values were determined according to the following formula: $t=(M_1-M_2)/s/(1/3+1/3)^{1/2}$. When

the determined t-value was greater than a t-value at P, which means the probability of significance level, of 0.05 or 0.01 in the t-distribution table with 4 degrees of freedom, the difference was judged to be found in the expression level of the gene between the two types of cells at $p<0.05$ or $p<0.01$, respectively. The tables also include the information of an increase (+) or decrease (-) in the expression level of a gene in the exposed cells in comparison with that of unexposed cells.

The expression levels of the following clones were elevated 4 or 24 hours after the ultraviolet irradiation: HEMBA1002475, HEMBA1004055, HEMBA1004669, HEMBA1006676, HEMBA1007085, HEMBB1002600, MAMMA1000284, MAMMA1000416, NT2RM1000039, NT2RM2000101, NT2RM2001696, NT2RM4002593, NT2RP2000459, NT2RP2001327, NT2RP2001420, NT2RP2002193, NT2RP2002208, NT2RP2003272, NT2RP2004013, NT2RP2004242, NT2RP3000109, NT2RP3000605, NT2RP3001730, NT2RP3002273, NT2RP3003290, NT2RP4000973, OVARC1000781, OVARC1001270, OVARC1001726, PLACE1000133, PLACE1001845, PLACE1004506, PLACE1005409, PLACE1005603, PLACE1006037, PLACE1006469, PLACE1008947, PLACE3000242, PLACE4000052, THYRO1000401, Y79AA1000784, and Y79AA1001781.

Industrial Applicability

The present invention provides genes associated with stomach cancer. The genes associated with stomach cancer of this invention are genes, which expression levels have been found to change specifically in stomach cancer. Therefore, it is highly possible that the current diagnosis and treatment of stomach cancer are drastically modernized by the finding. Screening of stomach cancer has been currently performed against normal healthy individuals who have reached a predetermined age as subjects mainly by imaging using endoscope, X-ray examination, etc. If a highly specific tumor marker of stomach cancer would be available, it may enable an early diagnosis of stomach cancer with serum, and is expected to improve the detection rate of stomach cancer at an early stage by using the marker alone or by combining it with conventional methods. Furthermore, it becomes

possible to predict the presence of minute metastatic focus, which is hardly detected by imaging, using a metastasis marker, and the prognosis prior to the treatment using a prognostic marker.

In addition, since the genes of this invention are closely associated with carcinogenesis and malignancy of gastric tissues, these genes and proteins encoded by the genes are useful as target molecules of cancer treatment. Detection of compounds that control the functions of these genes and proteins can lead to the development of anti-cancer agents effective to progressive cancers.

Furthermore, the present invention provides genes, that are specifically expressed in the high metastatic capability cell strain, OCUM-2MD3. Genes based on this invention and proteins encoded by the genes are closely associated with peritoneal metastasis of scirrhous stomach cancer. Therefore, it can be predicted that a cancer of a patient is liable to cause peritoneal metastasis when the gene or protein is detected in the humor or excised cancer tissues of the patient. Thus, the present invention can be used to predict the malignancy grade of scirrhous stomach cancer.

On the other hand, it is highly likely that the genes of this invention, or proteins encoded by the genes play important roles in peritoneal metastasis of cancer cells. Therefore, inhibition of functions of these genes and proteins may enable prophylaxis or suppression of peritoneal metastasis. That is, the present invention can be used for screening compounds useful for prophylaxis or suppression of peritoneal metastasis. Since the proteins of the present invention are considered to play important roles in peritoneal metastasis, these proteins are important targets for producing drugs.